

UNIVERSIDADE FEDERAL DO PARANÁ

MARCOS BENEDITO SCHIMALSKI

CLASSIFICAÇÃO DA COBERTURA VEGETAL A PARTIR DOS DADOS BRUTOS  
ALS EMPREGANDO MÉTODOS ESTATÍSTICOS DE RECONHECIMENTO DE  
PADRÕES

CURITIBA  
2009

MARCOS BENEDITO SCHIMALSKI

CLASSIFICAÇÃO DA COBERTURA VEGETAL A PARTIR D  
OS DADOS BRUTOS ALS EMPREGANDO MÉTODOS ESTATÍSTICOS DE  
RECONHECIMENTO DE PADRÕES

Tese apresentada ao Curso de Pós-Graduação  
em Ciências Geodésicas, Setor de Ciências da  
Terra, Universidade Federal do Paraná, como  
requisito parcial à obtenção do título de Doutor em  
Ciências Geodésicas.

Orientadores: Prof. Dr. **Edson A. Mitishita**  
Prof. Dr. **Anselmo Chaves Neto**

CURITIBA  
2009



## TERMO DE APROVAÇÃO

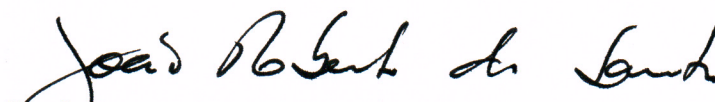
MARCOS BENEDITO SCHIMALSKI

### "CLASSIFICAÇÃO DA COBERTURA VEGETAL A PARTIR DOS DADOS BRUTOS ALS EMPREGANDO MÉTODOS ESTATÍSTICOS DE RECONHECIMENTO DE PADRÕES"

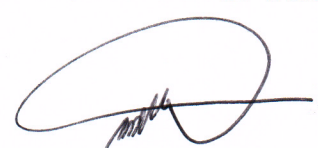
Tese nº 67 aprovada como requisito parcial do grau de Doutor no Curso de Pós-Graduação em Ciências Geodésicas, Setor de Ciências da Terra da Universidade Federal do Paraná, pela seguinte banca examinadora:

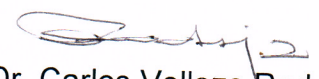
Orientadora:


  
Prof. Dr. Edson Aparecido Mitishita  
Departamento de Geomática, UFPR

  
Prof. Dr. João Roberto dos Santos  
Instituto Nacional de Pesquisas Espaciais, INPE

  
Prof. Dr. Attilio Antonio Disperati  
Universidade Estadual do Centro Oeste, UNICENTRO

  
Prof. Dr. Inácio Andruski Guimarães  
Universidade Tecnológica Federal do Paraná, UTFPR

  
Prof. Dr. Carlos Vellozo Roderjan  
Departamento de Ciências Florestais, UFPR

  
Prof. Dr. Quintino Dalmolin  
Departamento de Geomática, UFPR

Curitiba, 10 de dezembro de 2009.

## **DEDICATÓRIA**

Ao criador de todas as coisas visíveis e invisíveis e que sempre prestou auxílio do início até a conclusão deste trabalho: Deus.

À minha esposa Leila, com extrema gratidão e amor, pelo incentivo incondicional nos diversos momentos de dificuldades e preocupações vivenciados ao longo desta caminhada.

Ao meu filho João Victor, o sorriso em seus lábios e o brilho de seus olhos são sempre alento ao seu pai e serviram de estímulo para a conclusão desta tese.

Ao meu pai Paulo e minha mãe Antonia que sempre me incentivaram a trilhar a carreira acadêmica.

## AGRADECIMENTOS

A Deus, pela vida, bênção e proteção.

A Universidade do Contestado - Campus Canoinhas pelo apoio financeiro concedido através da bolsa de capacitação.

A empresa Aerosat pelo empréstimo do filme para a obtenção das fotografias aéreas.

A empresa Geobrasil pela confecção dos diapositivos empregados nas atividades de restituição fotogramétrica.

A empresa Fiducial pelo uso do *scanner* fotogramétrico empregado na obtenção das imagens em formato digital a partir do formato analógico.

Ao LACTEC pelo fornecimento dos dados *laser scanner*.

Ao fotogrametrista e amigo Romilton Pedrosa, pelas atividades de restituição fotogramétrica prestados gratuitamente.

Aos professores Christel Lingnau e Carlos Vellozo Roderjan por dedicarem tempo para o treinamento na fotointerpretação das formações vegetais.

Ao professor José Hilário Koehler (*in memoriam*) por sua ajuda nas atividades de reambulação e identificação das formações vegetais.

Ao professor Inácio Andruski Guimarães pela valiosa ajuda no estudo da Regressão Logística, bem como, a seção do algoritmo desenvolvido para esta ferramenta estatística.

Aos professores e funcionários do Curso de Pós-Graduação em Ciências Geodésicas, em destaque Verali Mônica Kleuser, pela ajuda prestada na realização deste trabalho e durante todo o doutorado.

Aos companheiros do Curso de Pós-Graduação em Ciências Geodésicas pelos momentos vivenciados juntos.

Ao Professor Anselmo Chaves Neto cujo auxílio e incentivo foram muito importantes para a implementação das ferramentas Estatísticas.

Finalmente, de forma especial, agradeço ao Professor Edson Aparecido Mitishita pela confiança e amizade a mim confiados, bem como, as suas sugestões e críticas, essenciais a obtenção do trabalho aqui apresentado.

*Date gloriam Deo,  
Deus tecum est in universis quae agis.*

Gênesis

## RESUMO

O setor florestal emprega informações referenciadas a um sistema geodésico para diferentes finalidades, visando orientar as suas operações de campo que são norteadas principalmente pelo código florestal brasileiro. A disponibilidade de bases cartográficas tridimensionais atualizadas é um fator chave para o sucesso das atividades florestais. Neste cenário, o *laser scanner* embarcado em aeronaves tem sido empregado para a derivação de produtos relacionados a morfologia do terreno mapeado, citando com exemplos as curvas de nível e o modelo digital do terreno (MDT). Havendo a necessidade de informações relativas à cobertura do solo, estas são geradas através de imagens digitais obtidas por sensores embarcados em plataformas aéreas ou espaciais. O objetivo desta pesquisa foi o desenvolvimento de metodologias que possibilitem a classificação da cobertura do solo com base nas informações pontuais oriundas da varredura *laser* e funções de reconhecimento de padrões da estatística multivariada. Foram realizados experimentos que classificaram os dados *laser* referentes a uma área rural, localizada nas proximidades da cidade de União da Vitória no estado do Paraná. Como cofatores de classificação foram selecionados a altura da cobertura vegetal, densidade de pontos, média e desvio padrão da altitude ortométrica e da intensidade de retorno do pulso *laser*. Para a obtenção do cofator relativo a altura das formações vegetais desenvolveu-se uma metodologia automatizada baseada na filtragem matemática dos pontos *laser* necessários a interpolação da superfície do terreno e obtenção das alturas médias destas formações. Para a classificação automática das classes relativas a cobertura do solo empregaram-se a Função Discriminante Linear de Fisher, a Função Discriminante Quadrática e a Regressão Logística Politômica com Camada Oculta. A exatidão do processo foi verificada através da comparação dos dados obtidos via as funções de classificação e os dados obtidos através da fotointerpretação manual das imagens aéreas do local. O melhor desempenho foi obtido através do emprego dos seis cofatores associados à Regressão Logística com Camada Oculta. Nesta configuração, verificou-se um índice de acerto médio de 75,9% na separação das classes temáticas.

Palavras-chave: *Laser scanning*, Classificação, Automatização, Função Discriminante Linear de Fisher, Função Discriminante Quadrática, Regressão Logística Politômica com camada oculta

## ABSTRACT

The forestry sector employs information referenced to a geodetic system for different purposes, in order to guide its field operations that are driven primarily by the Brazilian Forest Law. The availability of updated three-dimensional cartographic databases is a key factor for the success of forest activities. In this scenario, the laser scanner on board aircraft has been used for the derivation of products related to morphology of mapped terrain such as the contour lines and digital terrain model (DTM). If there is a need for information on land cover, these are generated from digital images obtained by sensors onboard airborne platforms or space. The purpose of this research was to develop methods that allow the classification of land cover based on accurate information coming from the laser scanning and functions of recognition of multivariate statistics. Experiments were carried out to classify the laser data relating to a rural area located near the city of Uniao da Vitoria, in the state of Parana. Vegetation height, density of points, average and standard deviation of the orthometric height and intensity of the return laser pulse were selected as cofactors of classification. To obtain the cofactor on the vegetation height, an automated methodology based on mathematical filtering of laser spots needed to interpolate the surface of the land and obtain the average heights of these formations was developed. For the automatic classification of classes for soil cover, the Fisher's Linear Discriminant Function, Quadratic Discriminant Function and Polytomous Logistic Regression with Hidden Layer were employed. The accuracy of the process was verified by comparing the data obtained by the ranking functions and data obtained through manual photo interpretation of aerial imagery of the site. The best performance was obtained through the use of six cofactors associated with Logistic Regression Hidden Layer. In this configuration, there was an average hit rate of 75.9% in the separation of the thematic classes.

Key words: Laser Scanner, Classification, Automation, Fisher's Linear Discriminant Function, Quadratic Discriminant Function, Polytomous Logistic Regression with Hidden Layer.

## LISTA DE ILUSTRAÇÕES

FIGURA 1 - EXEMPLO DE MDS, MDT E MDS <sub>N</sub> .....	23
FIGURA 2 - VARREDURA LASER .....	27
FIGURA 3 - PROJEÇÃO DO FOOTPRINT E VISUALIZAÇÃO DE MÚLTIPLOS RETORNOS.....	28
FIGURA 4 – COMPONENTES DO SISTEMA ALS .....	29
FIGURA 5 – DETERMINAÇÃO DO TEMPO DE RETARDO .....	30
FIGURA 6 – PRINCÍPIO DE MEDIÇÃO EM UM SISTEMA NO MODO CONTÍNUO ..	31
FIGURA 7 - FUNCIONAMENTO DO FILTRO PROGRESSIVO.....	34
FIGURA 8 - PROBLEMA CLÁSSICO DE RECONHECIMENTO DE PADRÕES .....	41
FIGURA 9 - CONFIGURAÇÃO DE GRUPOS .....	42
FIGURA 10 - MODELO DE REGRESSÃO LOGÍSTICO OCULTO COM VARIÁVEL RESPOSTA POLITÔMICA .....	64
FIGURA 11 – LOCALIZAÇÃO DA ÁREA DE ESTUDO.....	68
FIGURA 12 – LIMITE DA VARREDURA LASER E DA ÁREA DOS EXPERIMENTOS .....	71
FIGURA 13 – MODELO ESTEREOSCÓPICO 0203 .....	73
FIGURA 14 – MODELO ESTEREOSCÓPICO 0304 .....	74
FIGURA 15 – PAR FOTOGRÁFICO AJUSTADO SOB O ESTEREOSCÓPIO DE ESPELHOS.....	77
FIGURA 16 – MODELO 0203 COM AS CLASSES TEMÁTICAS DELIMITADAS.....	80
FIGURA 17 – DISTRIBUIÇÃO DOS PONTOS DE APOIO (GCP - <i>GROUND CONTROL POINT</i> ) NO MODELO FOTOGRAMÉTRICO .....	82
FIGURA 18 – DISTRIBUIÇÃO DOS PONTOS DE APOIO NAS FOTOGRAFIAS.....	83
FIGURA 19 – RASTREIO DE UM PONTO FOTO-IDENTIFICÁVEL.....	84
FIGURA 20 – DISPOSIÇÃO DA GRADE LASER E A ÁREA DOS EXPERIMENTOS	89
FIGURA 21 - EXEMPLO DE UMA CÉLULA DA GRADE DE 10 METROS .....	91
FIGURA 22 –CÁLCULO DA INCLINAÇÃO DO CONE.....	94

FIGURA 23 – VISUALIZAÇÃO DA VARIÁVEL ALTURA DA COBERTURA VEGETAL PARA UMA CÉLULA OCUPADA POR FLORESTA .....	97
FIGURA 24 – DELIMITAÇÃO DAS ÁREAS AMOSTRAIS .....	101
FIGURA 25 – ETAPAS DE CLASSIFICAÇÃO DA COBERTURA VEGETAL .....	114
FIGURA 26 - BASE CARTOGRÁFICA OBTIDA PELA RESTITUIÇÃO .....	118
FIGURA 27 - PERFIS LONGITUDINAIS PROVENIENTES DA RESTITUIÇÃO FOTOGRAMÉTRICA .....	121
FIGURA 28 - ANAGLIFO DA ÁREA 1.....	125
FIGURA 29 - ITERAÇÕES DO FILTRO MORFOLÓGICO PARA A ÁREA 1.....	126
FIGURA 30 - FOTOGRAMETRIA x FILTRO MORFOLÓGICO NA ÁREA 1 .....	127
FIGURA 31 - CURVAS DE NÍVEL PARA A ÁREA 1.....	127
FIGURA 32 - ANAGLIFO DA ÁREA 2.....	129
FIGURA 33 - ITERAÇÕES DO FILTRO MORFOLÓGICO PARA A ÁREA 2 .....	130
FIGURA 34 - COMPARAÇÃO VISUAL DAS GRADES REGULARES DA ÁREA 2 ..	131
FIGURA 35 - CURVAS DE NÍVEL PARA A ÁREA 2.....	131
FIGURA 36 - GRADES: FOTOGRAMETRIA x FILTRO MORFOLÓGICO.....	133
FIGURA 37 - MATRIZ DE DIFERENÇAS DE ALTURA PARA A ÁREA TOTAL.....	134
FIGURA 38 - ACERTO DA FDL COM COFADORES H E I .....	140
FIGURA 39 - ÍNDICES DE ACERTO DO 1º E 2º EXPERIMENTOS .....	145
FIGURA 40 - ÍNDICES DE ACERTO DO 1º, 2º E 3º EXPERIMENTO .....	152
FIGURA 41 - ÍNDICES DE ACERTO DO 3º E DO 4º EXPERIMENTO.....	159
FIGURA 42 - EFEITO DO AUMENTO DA RESOLUÇÃO GEOMÉTRICA DA GRADE REGULAR.....	160
FIGURA 43 - ÍNDICES DE ACERTO PARA O 3º, 4º E 5º EXPERIMENTO .....	165
FIGURA 44 - ÍNDICES DE ACERTO PARA O 3º E 6º EXPERIMENTO .....	170
FIGURA 45 - ÍNDICES DE ACERTO PARA O 3º, 6º E 7º EXPERIMENTO .....	176
FIGURA 46 - CLASSIFICAÇÃO DA COBERTURA DO SOLO OBTIDA COM A RLO - GRADE 10M .....	181



## LISTA DE TABELAS E QUADROS

QUADRO 1 - QUANTIDADE DE PONTOS LIDAR SELECIONADOS .....	71
QUADRO 2 - CHAVES DE FOTOINTERPRETAÇÃO SELECIONADAS .....	79
QUADRO 3 - CLASSES INSPECIONADAS EM CAMPO .....	86
TABELA 1 - RESULTADOS DO PROCESSAMENTO DO APOIO DE CAMPO.....	116
TABELA 2 - TOTALIZAÇÃO DAS CLASSES TEMÁTICAS .....	119
TABELA 3 - COMPARATIVO DE COORDENADAS LASER x FOTOGRAMETRIA .	123
TABELA 4 - CENTRÓIDES DAS CLASSES TEMÁTICAS PARA A GRADE 10M - VARIÁVEIS I E H.....	138
TABELA 5 - MATRIZ DE CONFUSÃO PARA A GRADE 10 M - VARIÁVEIS I E H ....	139
TABELA 6 - CENTRÓIDES DAS CLASSES TEMÁTICA PARA A GRADE 10M - VARIÁVEIS H,I E $\Delta h$ .....	144
TABELA 7 - MATRIZ DE CONFUSÃO PARA A GRADE 10 M - VARIÁVEIS H,I E $\Delta H$ .....	144
TABELA 8 - ÍNDICES DE ACERTO PARA AS CLASSES NOS DOIS EXPERIMENTOS.....	145
TABELA 9 - COEFICIENTES DAS FUNÇÕES DE DISCRIMINAÇÃO PARA A GRADE REGULAR DE 10 M EMPREGANDO TODAS AS VARIÁVEIS .....	150
TABELA 10 - CENTRÓIDES DAS CLASSES TEMÁTICA PARA A GRADE 10M - TODAS AS VARIÁVEIS .....	151
TABELA 11 - MATRIZ DE CONFUSÃO PARA A GRADE 10 M - TODAS AS VARIÁVEIS .....	151
TABELA 12 - ÍNDICES DE ACERTO DAS CLASSES PARA OS TRÊS EXPE- RIMENTOS .....	152
TABELA 13 - COEFICIENTES DAS FUNÇÕES DE DISCRIMINAÇÃO PARA A GRADE REGULAR DE 5 M EMPREGANDO TODAS AS VARIÁVEIS	157
TABELA 14 - CENTRÓIDES DAS CLASSES TEMÁTICAS PARA A GRADE 5M - TODAS AS VARIÁVEIS .....	157
TABELA 15 - MATRIZ DE CONFUSÃO PARA A GRADE 5 M - TODAS AS VARIÁVEIS .....	158

TABELA 16 - ÍNDICES DE ACERTO PARA AS CLASSES NO 3° E 4° EXPERIMENTO .....	158
TABELA 17 - COEFICIENTES DAS FUNÇÕES DE DISCRIMINAÇÃO PARA A GRADE REGULAR DE 1 M EMPREGANDO TODAS AS VARIÁVEIS	163
TABELA 18 - CENTRÓIDES DAS CLASSES TEMÁTICA PARA A GRADE 1M - TODAS AS VARIÁVEIS .....	164
TABELA 19 - MATRIZ DE CONFUSÃO PARA A GRADE 1 M - TODAS AS VARIÁVEIS .....	164
TABELA 20 - ÍNDICES DE ACERTO PARA AS CLASSES NO 3°, 4° E 5° EXPERIMENTO .....	165
TABELA 21 - MATRIZ DE CONFUSÃO PARA A GRADE 10 M - TODAS AS VARIÁ- VEIS - FUNÇÃO DISCRIMINANTE QUADRÁTICA.....	169
TABELA 22 - ÍNDICES DE ACERTO PARA AS CLASSES NO 3° E 6° EXPERIMENTO .....	170
TABELA 23 - COEFICIENTES DAS FUNÇÕES DE DISCRIMINAÇÃO PARA A GRADE 10 M - TODAS AS VARIÁVEIS - RLO .....	175
TABELA 24 - MATRIZ DE CONFUSÃO PARA A GRADE 10 M - TODAS AS VARIÁVEIS - RLO .....	175
TABELA 25 - ÍNDICES DE ACERTO PARA AS CLASSES NO 3° E 7° EXPERIMENTO .....	176

## SUMÁRIO

<b>1 INTRODUÇÃO</b>	14
1.1 CONSIDERAÇÕES GERAIS	14
1.2 OBJETIVOS	18
1.2.1 Objetivo Geral	18
1.2.2 Objetivos Específicos	18
1.3 - JUSTIFICATIVAS	19
<b>2 REVISÃO DE LITERATURA</b>	21
2.1 MODELO DIGITAL DO TERRENO	21
2.2 MODELO DIGITAL DE SUPERFÍCIE	22
2.3 FOTOGRAMETRIA E FOTOINTERPRETAÇÃO	24
2.4 SISTEMA LASER SCANNER	25
2.4.1 Conceitos Fundamentais	26
2.4 CLASSIFICAÇÃO DE DADOS LASER SCANNER	32
2.5.1 Métodos de Filtragem dos Dados LASER <i>Scanner</i>	32
2.5.2 Utilização de Dados LASER Scanner na Classificação de Tipologias Florestais	35
2.6 ESTATÍSTICA MULTIVARIADA	39
2.6.1 Reconhecimento de Padrões	40
2.6.2 Análise Discriminante	42
2.6.2.1 Função Discriminante Linear de Fisher para duas populações	43
2.6.2.2 Função Discriminante Linear de Fisher para diversas populações	48
2.6.2.3 Função Discriminante Quadrática	53
2.6.2.4 Função de Discriminação Logística	54
2.6.2.4.1 Modelo de Regressão Logístico Simples	56
2.6.2.4.2 Ajustamento do Modelo de Regressão Logístico Simples	57
2.6.2.4.3 Modelo de Regressão Logístico Múltiplo	59
2.6.2.4.4 Modelo de Regressão Logístico Múltiplo com resposta politômica	60
2.6.2.4.2 Modelo de Regressão Logístico Oculto	63
<b>3 MATERIAL E MÉTODOS</b>	68
3.1 LOCALIZAÇÃO DA ÁREA DE ESTUDO	68
3.2 MATERIAIS	69
3.3 METODOLOGIA DE EXECUÇÃO DA PESQUISA	72
3.3.1 Fotointerpretação das Fotografias Aéreas Verticais	72
3.3.2 Reambulação e Obtenção do Apoio Fotogramétrico	81
3.3.3 Referenciamento Cartográfico da Fotointerpretação	87
3.3.4 Criação das Grades Regulares para o Armazenamento dos Pontos LASER	88
3.3.5 Determinação do Modelo Digital do Terreno	91
3.3.6 Determinação da Altura da Cobertura Vegetal	96
3.3.7 Obtenção das Amostras Referentes à Cobertura Vegetal	97
3.3.8 Ajuste das Funções Discriminantes	102
3.3.8.1 Ajuste da Função Discriminante de Fisher	104
3.3.8.2 Ajuste da Função Discriminante Quadrática	106
3.3.9 Determinação da Cobertura Vegetal Através da Classificação das Grades Regulares	109
3.3.9.1 Classificação Baseada na Função Discriminante de Fisher	109
3.3.9.2 Classificação Baseada na Função Discriminante Quadrática	110

3.3.9.3 Classificação Baseada na RLO .....	110
3.3.9.4 Determinação e Análise dos Índices de Acerto das Funções Discriminantes .....	111
<b>4 RESULTADOS E DISCUSSÕES .....</b>	<b>115</b>
4.1 APOIO DE CAMPO.....	115
4.2 RESTITUIÇÃO FOTOGRAMÉTRICA E REAMBULAÇÃO.....	117
4.2.1 Restituição das Classes Temáticas .....	117
4.2.2 Determinação dos Perfis Longitudinais .....	120
4.2.3 Compatibilidade entre as coordenadas do mapa temático e a imagem de intensidade.....	122
4.3 DEFINIÇÃO DO MODELO DIGITAL DO TERRENO .....	124
4.3.1 Resultados da Filtragem para a Área 1 .....	124
4.3.2 Resultados da Filtragem para a Área 2 .....	128
4.3.3 Resultados da Filtragem para a Área Total .....	132
4.4 DETERMINAÇÃO DA COBERTURA VEGETAL ATRAVÉS DA CLASSIFICAÇÃO DAS GRADES REGULARES .....	135
4.4.1 Classificação Empregando a Função Discriminante de Fisher .....	137
4.4.1.1 Grade regular com resolução de 10 metros empregando-se as variáveis h e l .....	138
4.4.1.2 Grade regular com resolução de 10 metros empregando-se as variáveis H, l e altura Média na célula .....	143
4.4.1.3 Grade regular com resolução de 10 metros empregando-se todas as variáveis .....	150
4.4.1.4 Grade regular com resolução de 5 metros empregando-se todas as variáveis .....	156
4.4.1.5 Grade regular com resolução de 1 metro empregando-se todas as variáveis .....	163
4.4.2 Classificação Empregando a Função Discriminante Quadrática .....	169
4.4.3 Classificação Empregando a RLO .....	174
<b>5 CONCLUSÕES E RECOMENDAÇÕES .....</b>	<b>182</b>
<b>REFERÊNCIAS.....</b>	<b>187</b>
<b>APÊNDICE .....</b>	<b>193</b>

## 1 INTRODUÇÃO

### 1.1 CONSIDERAÇÕES GERAIS

As questões ambientais, principalmente aquelas relacionadas ao aquecimento global, têm ocupado lugar de destaque nos meios de comunicação, na sociedade e na esfera governamental. A pressão pela preservação e mesmo recuperação dos recursos naturais tem fortalecido a política e a legislação ambiental no Brasil. Neste contexto, o Código Florestal Brasileiro (Lei nº 4.771 de 15 de setembro de 1965) apresenta parâmetros norteadores às atividades no espaço físico, visando limitar as pressões sobre os recursos naturais além de subsidiar juridicamente as atividades de fiscalização e planejamento.

O Código Florestal Brasileiro define as áreas de preservação permanente (APP), cuja utilização é bastante restritiva. Neste caso, citam-se como exemplos, as formações naturais flúvio lacustres relacionadas à largura dos rios e áreas de encosta, nas quais a declividade seja superior a 45°.

Para verificar os parâmetros delimitadores do código florestal, devem ser empregadas informações atualizadas a respeito do espaço físico, de forma a orientar a ocupação humana, preservando os remanescentes naturais e possibilitando a execução de atividades para a recuperação das áreas degradadas.

Inserido nesta temática, o setor florestal necessita de informações referenciadas a um sistema cartográfico que possibilitem a verificação da legislação ambiental e a extração de parâmetros relacionados ao planejamento florestal. Usualmente, as informações necessárias ao planejamento são determinadas diretamente em campo, despendendo recursos na extração de observações relacionadas à estrutura vertical e horizontal da floresta, identificação e contagem de indivíduos, delimitação de atividades humanas, entre outras. Adiciona-se, ainda o fato que a região do planalto norte catarinense onde foi realizada a pesquisa, apresenta forte tradição em agropecuária e silvicultura, sendo que esta última ocupa uma área estimada de 600.000 ha.

O contexto apresentado evidencia a necessidade de informações tridimensionais sobre a superfície terrestre, que possibilitem verificar a observância

das leis ambientais e também derivem informações auxiliares às atividades florestais em campo, otimizando desta forma recursos.

As técnicas disponíveis para a obtenção de informações tridimensionais dividem-se naquelas que obtêm observações diretamente do terreno e as técnicas indiretas. As técnicas de mensuração indireta apresentam vantagens, principalmente devido à facilidade da obtenção das informações, comparativamente às técnicas de mensuração em campo. A geração de pares estereoscópicos associada às técnicas automáticas de geração do MDT (Modelo Digital do Terreno) são as mais comumente empregadas para a extração da altimetria (BALTSAVIAS (a), 1999). Contudo, a geração automática do Modelo Digital de Superfícies (MDS), que representa a cota superior dos objetos acima do terreno, é um grande problema especialmente em áreas florestadas.

Alternativa ao processo fotogramétrico, o ALS (*Airborne LASER Scanner* - LASER Scanner Aéreo) é uma inovação tecnológica bastante importante para o sensoriamento remoto, sendo uma tecnologia relativamente nova que possibilita a geração de coordenadas tridimensionais de uma nuvem de pontos sobre a superfície da Terra (DALMOLIN; SANTOS, 2003).

Outra denominação para o ALS é a sigla LIDAR (*Light Detection And Range* - Sistemas de Perfilamento a LASER). De acordo com MACHADO (2006), os sistemas LIDAR, integração de sistemas LASER (*Light Amplification by Stimulated Emission of Radiation*) scanner, GPS e INS (*Inertial Navigation System* - Sistema de Navegação Inercial), surgem como uma nova tecnologia de captura de informações sobre superfícies físicas, com um grande número de vantagens sobre o processo fotogramétrico convencional. A elevada automatização nos processos de coleta e processamento das informações, precisão, acurácia e alta densidade de pontos, além da rapidez, tornam os sistemas LIDAR inovadores.

Contudo, o processamento dos dados derivados pelo LIDAR carecem de metodologias que possibilitem a extração de feições classificadas em antrópicas (produzidas pelo homem, citando como exemplos, edificações e rodovias) e naturais (formações vegetais, rios e relevo como exemplos). Para possibilitar esta classificação, alguns autores sugerem a integração das observações LIDAR com fontes de dados espectrais, possibilitando desta maneira a obtenção de informações planimétricas. Habib *et al.* (2004) destacam que dados fotogramétricos e de LIDAR têm qualidades específicas que tornam uma solução ou outra preferencial para

certas aplicações. Uma desvantagem nas medições fotogramétricas é complementada por uma vantagem no sistema LIDAR, e vice-versa. As informações altimétricas obtidas pelo LIDAR tem qualidade muito superior quando comparadas com as informações planimétricas e o contrário pode ser verificado no processo fotogramétrico. Com isso, a integração entre informações espectrais e LIDAR apresenta benefícios interessantes.

No âmbito do processamento de dados espectrais estão disponíveis soluções comerciais e abertas que empregam conceitos de última geração relacionados com a segmentação de imagens, Redes Neurais Artificiais (RNA), lógica nebulosa (FUZZY), árvore de decisão e classificação orientada a objetos.

A segmentação de imagens se vale da morfologia matemática para auxiliar a classificação dos dados espectrais. Constitui-se em uma abordagem não-linear para análise espacial de estruturas e resolve de forma eficiente problemas de extração de informação, a partir da análise das estruturas geométricas dos alvos contidos nas imagens digitais (ALBERTI; DA SILVA, 2009). A morfologia matemática concentra seu esforço no estudo da estrutura geométrica das entidades presentes em uma imagem. Permite processar imagens com objetivos de realce, segmentação, detecção de bordas, esqueletização, afinamento e análise de formas entre outras.

Para Bernardes *et al.* (2009) a classificação orientada a objeto pressupõe uma abordagem intuitiva assemelhando-se aos métodos que os seres humanos utilizam para resolver problemas, ou seja, a simulação da percepção humana. A classificação orientada a objeto simula a interpretação visual através da modelagem do conhecimento pretendendo a identificação de feições orientada por padrões descritores como cor, textura, métrica, contexto e entropia, entre outras características. A utilização de técnicas de segmentação de imagens, manuseio hierárquico de múltiplas resoluções geométricas de imagens digitais, elaboração de redes semânticas e o emprego de lógica FUZZY são algumas das abordagens empregadas em conjuntos pelas soluções baseadas em contexto e orientado a objeto (BAATZ *et al.*, 2004).

Neste ponto deve ser destacado que o emprego de informações espectrais adicionais ao LIDAR, se traduz diretamente em um aumento de custos relacionados a obtenção e processamento destas informações. Destaca-se ainda que o processamento das informações espectrais não seja totalmente automatizado, o que

demanda grande quantidade de tempo para a extração das informações planimétricas a partir desta fonte de dados.

A proposta desta pesquisa é o desenvolvimento de metodologias automatizadas capazes de extrair as classes temáticas relacionadas com a cobertura do solo a partir dos dados brutos LIDAR, sem a necessidade de uma fonte adicional de informações. Como região de testes foi escolhida uma área de transição entre o ambiente urbano e o rural localizada nas proximidades de cidade de União da Vitória no estado do Paraná.

O reconhecimento de padrões foi empregado no desenvolvimento dos algoritmos de classificação, sendo escolhidas para os testes a Função Discriminante Linear de Fisher, a Função Quadrática e a Regressão Logística Politômica Oculta. Inicialmente, foi compilada a base cartográfica temática do local com o auxílio da fotointerpretação, restituição estereofotogramétrica e reambulação de campo. A partir desta base foram selecionadas e então obtidas amostras das classes temáticas a partir dos dados LIDAR, tornando possível o ajuste das funções de discriminação. O processo termina com a classificação dos dados LIDAR e determinação dos índices de acerto para cada função empregada.

Como cofatores das funções de discriminação escolheu-se: altitude e intensidade média, desvio padrão da altitude e da intensidade, densidade de pontos e altura da cobertura vegetal. Para a obtenção deste último cofator foi necessário desenvolver um filtro para a separação dos pontos LIDAR localizados na superfície topográfica e acima desta. O filtro morfológico empregado para esta finalidade foi proposto por ZAKSEK e PFEIFER (2006).

São apresentados a seguir os objetivos desta tese e as justificativas para a realização da pesquisa. Na seqüência apresenta-se a revisão da literatura consultada englobando os conceitos de LIDAR, Fotointerpretação, filtragem de dados provenientes da varredura LASER e reconhecimento de padrões.

A metodologia e materiais utilizados são apresentados e por último procede-se a análise dos resultados alcançados, finalizando-se o trabalho com a apresentação de conclusões e de recomendações.



## 1.2 OBJETIVOS

### 1.2.1 Objetivo Geral

Desenvolver metodologias que possibilitem a classificação da cobertura vegetal, utilizando as informações brutas oriundas da varredura aérea do LIDAR, mediante o emprego de funções discriminantes.

### 1.2.2 Objetivos Específicos

- Adequar um filtro matemático que possibilite a separação dos pontos LASER localizados na superfície topográfica;
- Comparar os resultados da filtragem matemática com a perfilagem fotogramétrica;
- Determinar as variáveis e funções discriminantes para a classificação da cobertura vegetal;
- Empregar as Funções Discriminantes: Fisher, Quadrática e Logística visando à classificação dos dados LIDAR de acordo com o uso do solo detectado pela Fotointerpretação e Reambulação de Campo;
- Analisar os índices de acerto das funções de discriminação tomando por referência a classificação da cobertura vegetal obtidos com a restituição estereofotogramétrica e reambulação de campo;
- Analisar o efeito da generalização de informações quando da utilização de grades regulares representativas da nuvem de pontos;
- Propor uma metodologia adequada à classificação baseada nas Funções Discriminantes analisadas.

### 1.3 - JUSTIFICATIVAS

A resolução geométrica das imagens provenientes dos sensores orbitais melhorou consideravelmente, após a liberação da tecnologia em 1994, pelo governo norte americano, sendo que a partir disto empresas privadas entraram no mercado de imageamento da superfície terrestre, tornando-o mais competitivo. Os sensores de alta resolução espacial têm revolucionado os paradigmas da interpretação de imagens, tornando viável a identificação de objetos que não eram visíveis com as tecnologias de imageamento até então disponíveis (BOTELHO, 2007).

O setor florestal tem empregado as imagens de alta resolução espacial para a obtenção de informações que possam auxiliar as suas atividades de campo, bem como, a verificação da legislação ambiental. São exemplos disto, a extração de informações referentes à área, perímetro, comprimentos, distribuição espacial de árvores e identificação de espécies florestais entre outras informações de natureza bidimensional. Havendo a necessidade de informações altimétricas, as mesmas são derivadas a partir de pares estereoscópicos e alternativamente através de recobrimentos aéreos empregando os sistemas LIDAR.

A varredura LASER obtém uma nuvem de pontos tridimensionais distribuídos irregularmente, armazenados em arquivos digitais e que despendem um esforço elevado para seu processamento. Alia-se o fato que estes pontos não são categorizados, isto é, separados em classes temáticas no instante de sua mensuração, sendo necessária a aplicação de técnicas especializadas para a extração das informações cartográficas.

Normalmente estas classes temáticas são derivadas com o auxílio de fontes de dados adicionais provenientes de imagens geradas por sensores ópticos ou ainda a partir de imagens derivadas da nuvem de pontos geradas pelo LIDAR, citando-se como exemplo, a imagem de intensidade. Apesar das técnicas de processamento de imagens estarem em um estágio de desenvolvimento avançado, a classificação não ocorre automaticamente sendo necessária a intervenção humana em diversas etapas do processo de classificação, além das inspeções de campo para validação dos resultados. Desta forma os produtos cartográficos acabam tendo o seu custo e prazos de obtenção elevados.

Neste cenário, as funções discriminantes podem auxiliar no melhor aproveitamento das informações provenientes da nuvem de pontos geradas pelo LIDAR, possibilitando a extração das classes temáticas sem a necessidade de fontes de dados adicionais. A partir de amostras referentes às classes temáticas pretendidas é possível com auxílio destas funções classificar a cobertura vegetal da superfície que foi recoberta pela varredura LASER.

Pesquisas recentes sobre a acurácia nas observações geradas pelo LIDAR, apontam para a necessidade da calibração das mensurações obtidas por estes sistemas. Um ponto obtido através do ALS é um produto composto de muitas medidas e eventos: georreferenciamento direto, determinação do ângulo do feixe LASER, emissão do feixe, interações do LASER (com a atmosfera, com a superfície terrestre e com os objetos acima desta), retorno do feixe, eventualmente detecção do sinal de retorno e medição do tempo de viagem do feixe. Todos estes processos têm um impacto sobre as coordenadas do ponto obtido. A classificação da cobertura vegetal passa a ter importância nesta temática, sendo informação útil para a atenuação do erro sistemático produzido pela vegetação baixa que afeta as mensurações do LIDAR.

O tipo de vegetação tem influência sobre o retorno do feixe, tendendo a distorcer a sua forma, e geralmente causando erros maiores. A vegetação de menor porte causa um deslocamento sistemático das mensurações (PFEIFER *et al.*, 2004). A classificação da cobertura vegetal pode identificar os locais ocupados por vegetação de menor altura que não é removida satisfatoriamente pelos filtros disponíveis, possibilitando desta forma a obtenção de medidas verticais mais acuradas e isentas deste tipo de erro.

Concluindo esta exposição ressalta-se ainda que a metodologia proposta trabalha diretamente com as observações brutas geradas pelo ALS, evitando-se o processamento de imagens digitais, que normalmente, são derivadas a partir da nuvem de pontos LASER. A automatização proposta, impacta na velocidade da obtenção da informação, diminuindo a intervenção humana no processo e conseqüentemente os erros.

## 2 REVISÃO DE LITERATURA

### 2.1 MODELO DIGITAL DO TERRENO

O terreno ou relevo pode ser definido como o conjunto de formas da superfície terrestre, resultado da ação de agentes endógenos (vulcanismo e tectonismo) e exógenos (intemperismo e humano). Outra definição é como o conjunto de características físicas de um trecho da superfície terrestre no qual são verificados relacionamentos entre elementos naturais e culturais (criados pela intervenção humana). A representação das irregularidades do terreno pode ser obtida através da modelagem baseada na amostragem desta superfície (ANDREWS, 1996).

O conceito de MDT foi aplicado pela primeira vez para projetos de estradas no *Massachusetts Institute of Technology* (MIT), em 1950. O seu aperfeiçoamento foi impulsionado pelo desenvolvimento da informática, a qual não só reduziu as dificuldades das rotinas de cálculo, como também ampliou suas áreas de aplicação (CAMPOS, 1994).

De acordo com Felgueiras (2001) um MDT é uma representação matemática computacional do relevo a partir de amostragem na superfície terrestre. Para a representação de uma superfície real no computador é indispensável a elaboração e criação de um modelo digital, que pode estar representado por equações analíticas ou uma rede (grade) de pontos, de modo a transmitir ao usuário as características espaciais do terreno.

Para gerar um MDT são necessárias as seguintes etapas (NAMIKAWA *et al.*, 2003):

- aquisição das amostras ou amostragem: compreende a aquisição de um conjunto de amostras representativas do fenômeno de interesse. Geralmente essas amostras estão representadas por curvas de isovalores e pontos tridimensionais;
- modelagem: envolve a criação de estruturas de dados e a definição de superfícies de ajuste com o objetivo de se obter uma representação contínua do fenômeno a partir das amostras. Essas estruturas são definidas de forma a possibilitar uma manipulação conveniente e eficiente dos modelos pelos algoritmos

de análise contidos no Sistema de Informações Geográficas (GIS). Os modelos mais utilizados são os modelos de grade regular retangular e os modelos de grade irregular triangular;

- Aplicações: são procedimentos de análise executados sobre os modelos digitais. As aplicações podem ser qualitativas, tais como a visualização do modelo usando-se projeções geométricas planares ou quantitativas tais como cálculos de volumes e geração de mapas de declividades.

Ainda de acordo com Namikawa *et al.* (2003) um Modelo Digital do Terreno poderá ter como representações computacionais:

- Pontos tridimensionais: são pontos com dados da posição espacial no plano xy e mais a altitude (z). Esta representação é empregada, por exemplo, para definirem-se pontos de máximos e de mínimos em cartas topográficas;
- Isolinhas: são linhas formadas por um conjunto ordenado de pontos bidimensionais definidos no plano xy. Cada isolinha contém, além dos pontos que define sua geometria a altitude;
- Grade irregular triangular (Triangular Irregular Network - TIN): é uma estrutura poliédrica cujos elementos básicos são triângulos. Os vértices dos triângulos pertencem ao conjunto de amostras do modelo;
- Grade regular retangular: é uma estrutura matricial que contém pontos tridimensionais regularmente espaçados no plano xy.

## 2.2 MODELO DIGITAL DE SUPERFÍCIE

O Modelo Digital de Superfície (MDS) se distingue do MDT por ser uma descrição do terreno e dos objetos acima deste. O MDS é constituído por uma nuvem de pontos tridimensionais que pode ser obtida por processo estéreo-fotogramétrico, cujas informações procedem de imagens digitais aéreas ou orbitais de alta resolução geométrica e ainda obtidas com auxílio de um LASER *Scanner* Aerotransportado (ALS).

Araki (2005) define o MDS como a representação computacional de todas as feições existentes acima do terreno, bem como a própria representação da

superfície (MDT). A diferença entre o MDT e o MDS resulta no Modelo Digital de Superfície Normalizado (MDSn).

O MDS de áreas urbanas é amplamente empregado em um crescente número de aplicações, entre estas, na produção de ortoimagens digitais, modelagem tridimensional de cidades e de edificações.

Na área florestal, LEE *et al.* (2004) propõe a criação de MDS de árvores isoladas e de áreas de floresta com o objetivo de monitorar a dinâmica do crescimento vegetal. A partir dos dados provenientes de uma varredura LIDAR associados a observações de campo foi possível determinar a quantidade de carbono absorvido da atmosfera pelas formações vegetais, bem como, gerar mapas que permitem uma melhor quantificação da recuperação estrutural da floresta e de suas propriedades, tais como folhagem, ramos e distribuição de troncos.

Na figura 1 são mostrados exemplos destas representações para uma área coberta por floresta natural. A figura 1(a) apresenta o MDS, a figura 1 (b) um MDT e a figura 1(c) o MDSn.

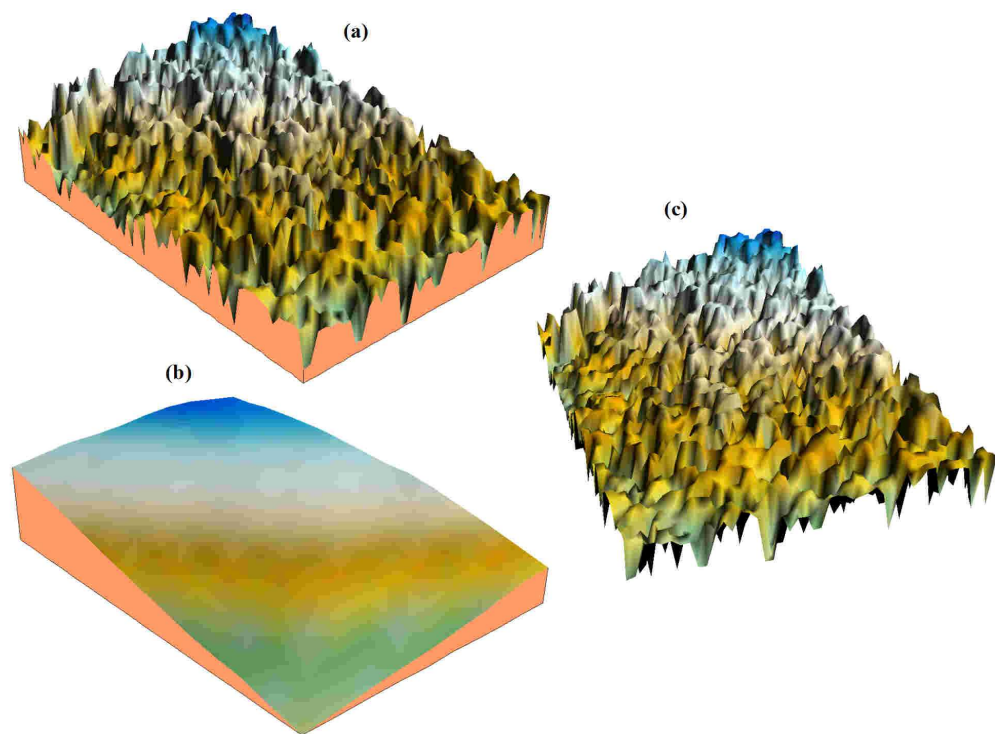


FIGURA 1 - EXEMPLO DE MDS, MDT E MDSn.

## 2.3 FOTOGRAMETRIA E FOTOINTERPRETAÇÃO

Nesta pesquisa, a fotogrametria e a fotointerpretação serviram como ferramentas auxiliares na obtenção da verdade de campo, visto que a outra alternativa seria a realização de inventários e levantamentos florestais a campo.

Segundo Andrade (2003) fotogrametria é a ciência e tecnologia de obter informações confiáveis através de processos de registro, interpretação e mensuração de imagens. O maior campo de aplicação da fotogrametria é a elaboração de mapas, auxiliada por outras ciências como a Geodésia e Cartografia. As imagens fotográficas são empregadas para o posicionamento de pontos da superfície terrestre e também para mapear objetos fotografados, tais como rede de drenagem, florestas, culturas, rede viária, feições geológicas e tipos de solo, entre outros.

A Fotointerpretação pode ser definida como o ato de examinar imagens fotográficas com o fim de identificar objetos e determinar seus significados (WOLF, 2000). Para Teng (1997) a Fotointerpretação é a inferência de informação a partir de dados obtidos através da observação de um meio fotográfico, com ou sem a ajuda de equipamento.

Disperati *et al.* (2007) ponderam ainda que qualquer tentativa de conceituar o termo Fotointerpretação, deverá envolver as seguintes termos: arte, ciência, meio fotográfico, observação, inferência, equipamento, julgamento, subjetividade, identificação de objetos e determinação de significados.

As pessoas sem experiência em Fotointerpretação podem apresentar dificuldades iniciais de interpretar uma fotografia aérea vertical. Estas dificuldades poderão estar relacionadas:

- à perspectiva vertical das fotografias aéreas, diferente da atividade humana normal que utiliza a visão em perspectiva horizontal;
- desconhecimento do local registrado pela fotografia aérea;
- à escala das fotografias, a qual a maioria das pessoas não considera habitualmente;
- em algumas situações especiais, nas quais a energia registrada não corresponde aos comprimentos de onda do visível, por exemplo, o infravermelho.

A Fotointerpretação é uma técnica e não uma ciência, como a Geologia, a Agronomia, entre outras. Enquanto técnica, ela precisa estar ligada a, pelo menos, um ramo profissional. Ela engloba um conjunto de técnicas que se aplicam a vários setores, como exemplos, a ecologia, florestal, aos estudos de zonas urbanas entre outros.

Para Loch (1993) a fotointerpretação depende diretamente do foto-intérprete, que pode dispor de técnicas muito desenvolvidas, fazendo com que a fotointerpretação seja quase exata e capaz de obter o maior número de informações da foto. O fot- intérprete deve apresentar como características fundamentais:

- Acuidade visual: é a capacidade do indivíduo de separar detalhes identificáveis nos objetos visíveis, fato que depende do poder de resolução do olho;
- Acuidade mental: depende da experiência do foto-intérprete e inclui bom senso, imaginação e perícia;
- Paciência e adaptabilidade.

## 2.4 SISTEMA LASER SCANNER

De acordo com Machado (2006) a concepção do LASER surgiu por volta de 1905, quando Albert Einstein propôs que poderiam ser obtidas emissões de luz estimuladas. Fontes de luz como velas, substâncias fluorescentes, e lâmpadas, emitem quantidades de energia, chamadas fótons, quando seus átomos são excitados pela energia. Einstein sugeriu que os átomos poderiam ser estimulados para emitir fótons. Esta emissão poderia produzir um tipo de luz altamente concentrada, brilhante, e poderosa, o que poderia ser utilizado em diversas atividades. Contudo, isto não contribuiu para a criação de um LASER devido às limitações tecnológicas da época. Esta concepção tornou-se real quando Charles Townes obteve em 1954, o que ele denominou de MASER (*Microwave Amplification by Stimulated Emission of Radiation* - amplificação de microondas pela emissão estimulada de radiação). Este protótipo estimulava os fótons na frequência das microondas, transformando o MASER numa fonte amplificadora de microondas.

Com esta inovação vários cientistas começaram a desenvolver pesquisas visando a construção de um MASER óptico, que posteriormente recebeu a



denominação de LASER (*Light Amplification by Stimulated Emission of Radiation* - amplificação da luz pela emissão estimulada de radiação).

O primeiro LASER óptico foi concebido por Theodore Maiman, físico norte-americano, que em 1960 conseguiu montar um LASER com funcionamento baseado em um rubi. O feixe LASER distingui-se pois é coerente (sincronismo de fases). Cada átomo é coerente com a luz que atravessa o material. Com tais características de coerência, alta intensidade, e precisão direcional, entre outras, fez com que o LASER tenha sido empregado em várias aplicações de diferentes setores.

Entre as aplicações desenvolvidas cita-se a medição de distâncias com o LASER operado próximo ao solo (terrestre) ou embarcado em uma aeronave.

#### 2.4.1 Conceitos Fundamentais

Para Centeno (2004) o sistema de varredura LASER é um sistema ativo usado para medir a distância entre o sistema sensor e a superfície, e a partir da integração com as informações coletas pelo Sistema de Posicionamento Global (GPS) e do Sistema de Medição Inercial (IMU) determinar a posição espacial ou geodésica do local onde o LASER foi refletido. O funcionamento do sistema baseia-se na utilização de um feixe LASER emitido na direção dos objetos. Ao atingir a superfície parte deste sinal é refletido para o sensor no qual este retorno é medido, juntamente com a determinação da diferença de tempo decorrido entre a emissão do pulso e recepção do retorno. A distância entre a feição imageada e o sensor pode ser então calculada. Outro parâmetro medido é a intensidade do retorno de sinal que fornece informações a respeito das características da superfície naquele ponto. Outras denominações para o sistema são: LIDAR (*Light Detection And Ranging*) ou LADAR (*LASER Detection and Ranging*).

A frequência do feixe LASER situa-se no intervalo de 500 a 1500 nm, com valores típicos de 1040 a 1060 nm (MACHADO, 2006). Estas frequências não são absorvidas pelo terreno, mas sim refletidas. As formações vegetais e nuvens apresentam também a capacidade de refletir estes comprimentos de onda, sendo que não é possível realizar a varredura com nuvens se interpondo entre a aeronave e o terreno. Todavia, estes sistemas apresentam mais condições operacionais do

que os sensores fotogramétricos: eles não sofrem com a falta de iluminação ou com sombras. Como seus sensores são ativos, é possível utilizá-los durante qualquer período do dia ou estação climática.

Quando comparado com as técnicas de microondas, os LASERs têm a vantagem que a medição das distâncias dos pontos sobre o solo é realizada com pulsos de energia concentrados em um curto intervalo de tempo (AXELSSON, 1999). O processamento básico dos dados LASER envolve a separação de pontos no terreno e dos objetos, de forma a possibilitar a geração do MDT e do MDS. Ainda segundo este autor, um sistema LASER *scanner* produz dados que podem ser caracterizados como sub-randômicos e que estão distribuídos em uma nuvem de pontos posicionados tridimensionalmente. Os dados referentes a elevação podem ser adquiridos com diferentes atributos, citando como exemplo, a densidade de pontos.

A densidade dos pontos depende da altura de voo e também dos parâmetros do sistema, como a velocidade da plataforma, FOV (*Field of View*) e frequência da amostragem. A amplitude de registro fornece informações a respeito da refletância de cada ponto imageado, possibilitando a geração de uma imagem de intensidade. A Figura 2 mostra a operação da varredura LASER.

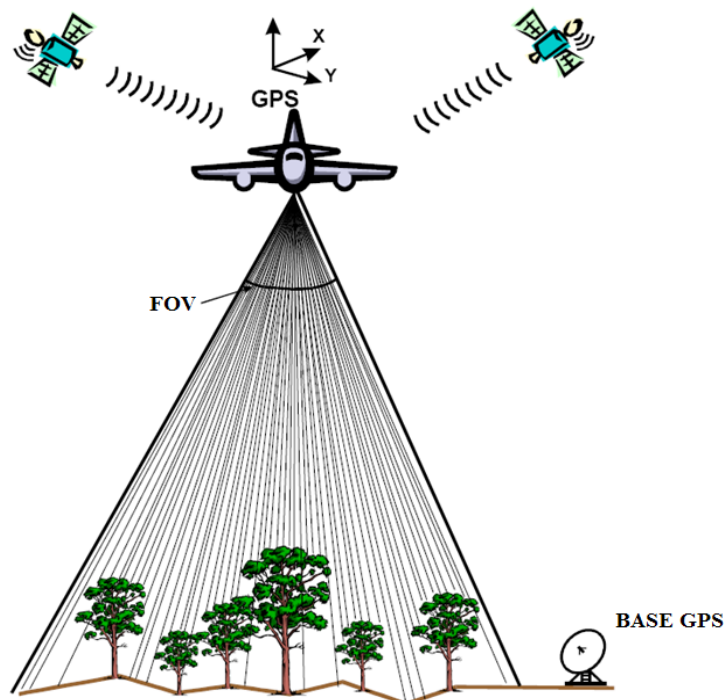


FIGURA 2 - VARREDURA LASER  
FONTE: Hyyppa *et al.* (2000)

Outra característica do feixe LASER é sua baixa divergência, o que torna a sua trajetória similar ao conceito geométrico de linha reta, mesmo quando a projeção de seu cone, denominada *footprint*, esteja com o diâmetro de 30 a 150 centímetros. Em função do *footprint*, poderá ocorrer mais de uma reflexão para o mesmo objeto. Estas múltiplas reflexões poderão ser empregadas para determinar a altura de um objeto.

A Figura 3 apresenta o *footprint* e a visualização de três reflexões ou ecos.

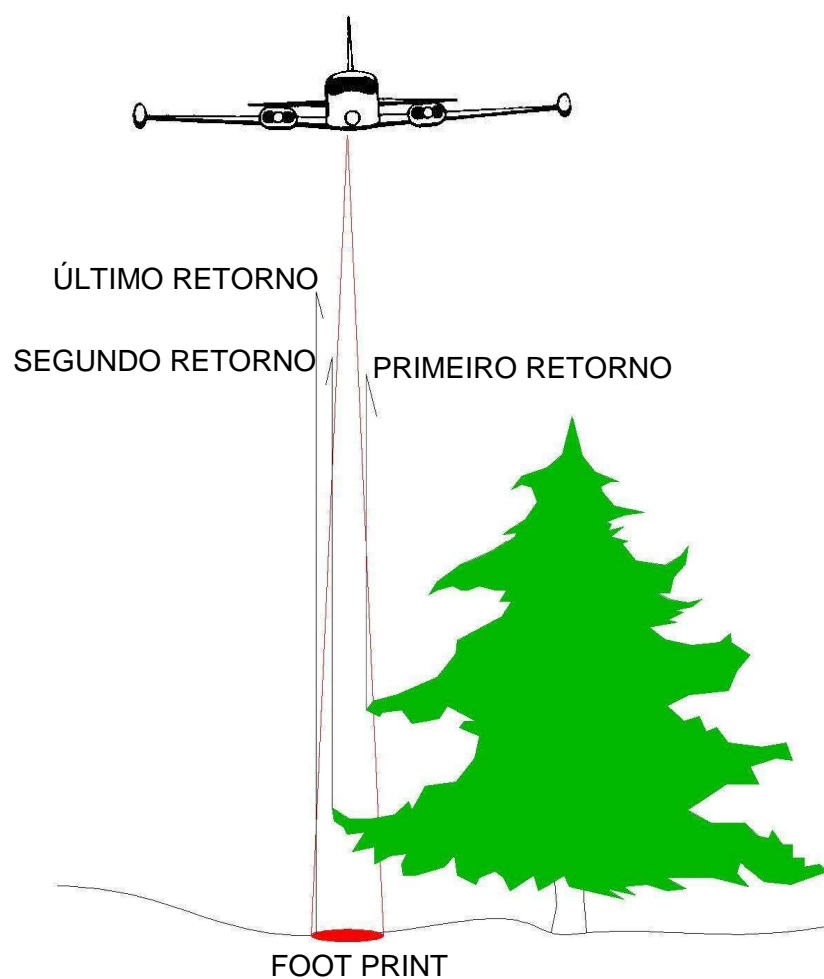


FIGURA 3 - PROJEÇÃO DO FOOTPRINT E VISUALIZAÇÃO DE MÚLTIPLOS RETORNOS

Com relação ao sistema LASER *scanner*, este pode apresentar três componentes principais: a unidade de medição LASER, sistema de varredura óptico-mecânico e uma unidade de registro/medição de pulso (WEHR;LOHR, 1998). A Figura 4 mostra um sistema ALS com seus elementos principais.

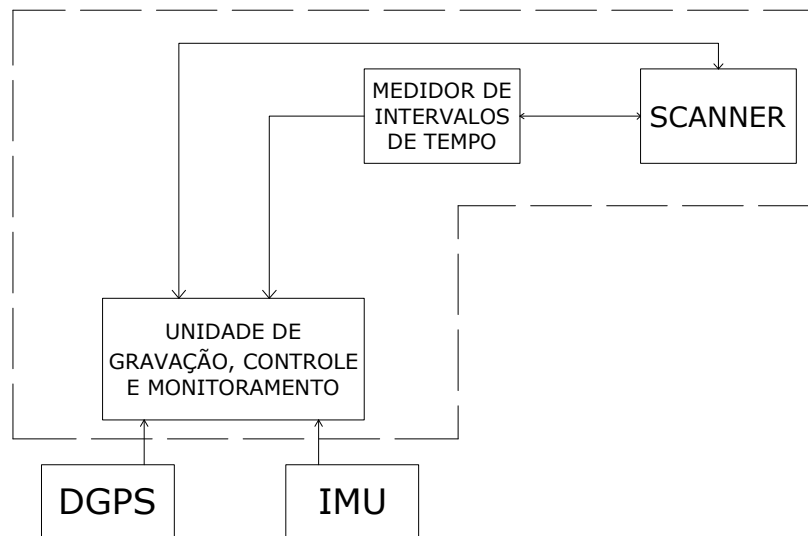


FIGURA 4 – COMPONENTES DO SISTEMA ALS

FONTE: Adaptado de Wehr e Lohr (1998)

A IMU (*Inertial Measurement Unit*) é encarregada de calcular a inclinação do sensor nas três direções (*roll* ( $\varpi$ ), *pitch* ( $\varphi$ ) e *heading* ( $\kappa$ )) e operando em conjunto com o DGPS (*Diferencial Global Positioning System*) possibilita a determinação da posição do sensor no momento da medição. A unidade de registro e de medição é responsável pela determinação da distância entre o sensor e a superfície imageada. Esta unidade pode operar de duas maneiras: com o sistema de pulsos e com o sistema de ondas contínuas (CENTENO, 2004). O sistema mais empregado nas soluções comerciais é o de pulso, que funciona com o princípio de emissão e recepção de pulsos curtos de energia. O valor de distância é calculado em função do tempo transcorrido entre a emissão e a recepção (sensor/objeto).

Ainda de acordo com Dalmolin e Santos (2003) no sistema de pulso, o tempo de retardo ( $dt$ ) é medido em relação a um determinado ponto do pulso. A diferença entre a emissão e o registro de retorno é determinada pela análise do retardo do pulso captado comparado com o emitido. A Figura 5 possibilita verificar a determinação do valor de  $dt$ .

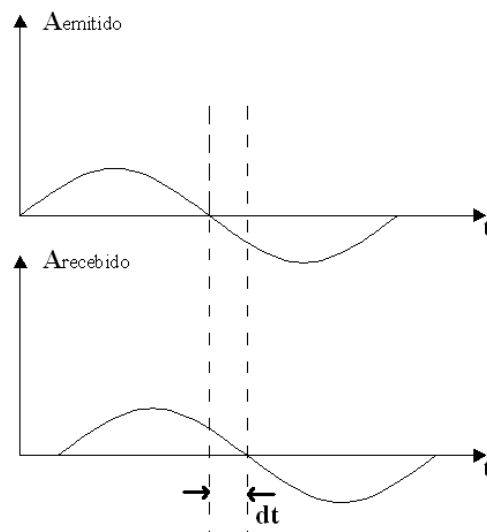


FIGURA 5 – DETERMINAÇÃO DO TEMPO DE RETARDO

FONTE: Dalmolin e Santos (2003).

Na Figura 5,  $A_{emitido}$  e  $A_{recebido}$  são respectivamente a amplitude do sinal. O cálculo da distância entre o sensor e o ponto imageado na superfície é

$$D = \frac{c \cdot dt}{2} \quad (1)$$

Sendo:

$D$  – Distância entre o sensor e a superfície imageada;  
 $dt$  - Diferença de tempo entre a emissão e a recepção;  
 $c$  - Velocidade da luz no vácuo.

Para um sistema operante no modo contínuo, o intervalo de tempo é calculado entre a defasagem da onda de retorno e a onda emitida. A Figura 6 mostra esta situação.

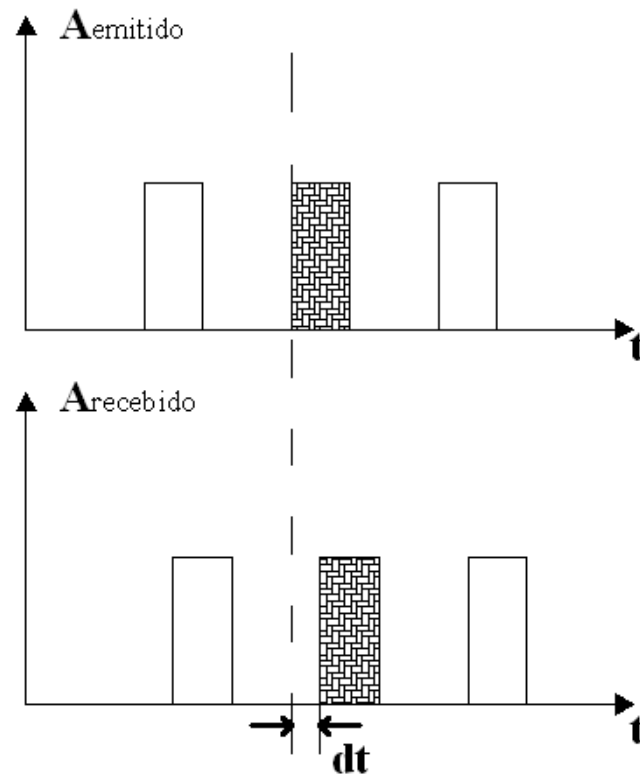


FIGURA 6 – PRINCÍPIO DE MEDIÇÃO EM UM SISTEMA NO MODO CONTÍNUO  
 FONTE: Adaptado de WEHR e LOHR (1999)

O cálculo de distância neste caso é determinado com o cálculo do intervalo entre a emissão e a recepção, sendo este intervalo diretamente proporcional à diferença de fase ( $\Phi$ ). Sendo conhecido o período da onda ( $P$ ) e o número de ciclos inteiros ( $n$ ), é possível calcular o intervalo entre a emissão e a recepção através da seguinte fórmula (WEHR;LOHR,1999):

$$dt = \frac{\Phi}{2 * \pi} * P + n * P \quad (2)$$

Em um LASER scanner as aberturas de transmissão e recepção, normalmente entre 8 e 15 centímetros, são montadas de tal forma que o feixe transmitido intercepte um ponto sobre o terreno e dentro do FOV (Campo de visão). A Figura 2 apresenta o FOV no instante da varredura.

Já o campo de visão instantâneo (IFOV) é definido pela divergência do feixe LASER, que varia entre 0,3 a 2  $\mu$ rad. O limite do IFOV é determinado pela difração da luz, que determina a luminosidade do local imageado. Desta forma, o IFOV é

função da abertura de transmissão ( $E$ ) e o comprimento do LASER ( $\lambda$ ) (BALTSAVIAS(b), 1999). Para a obtenção de um feixe LASER espacialmente coerente, a difração limite é determinada por:

$$\text{IFOV}_{\text{diff}} = 2,44 \cdot \frac{\lambda}{E} \quad (3)$$

Para se obter um IFOV compatível, o feixe LASER tem que ser movido cruzando a linha de vôo de forma a cobrir a área requerida para a execução do levantamento.

Os LASER *scanners* podem ser classificados de acordo com o tipo de varredura, o que é função dos espelhos de varredura óptico-mecânico empregados. São três os tipos de espelhos empregados na varredura: oscilador, Palmer e polígono de rotação. Um detalhamento sobre o funcionamento destes sistemas de varredura pode ser verificado em Dalmolin e Santos (2003).

## 2.5 CLASSIFICAÇÃO DE DADOS LASER SCANNER

Comumente os dados provenientes da varredura a LASER são utilizados para a extração do MDT e do MDS. Porém, a capacidade de registrar a intensidade que corresponde às características de refletância da superfície imageada, habilita o ALS a fornecer informações qualitativas, tal como a separação das classes de uso do solo. Neste tópico são abordadas metodologias para a classificação da nuvem de pontos LASER no que diz respeito à separação de pontos no solo e acima deste (filtragem), bem como a classificação de acordo com classes temáticas como, por exemplo, vegetação e água.

### 2.5.1 Métodos de Filtragem dos Dados LASER *Scanner*

O ALS fornece uma nuvem de pontos (com coordenadas geodésicas

tridimensionais) medidos sobre objetos localizados entre o *laser scanner* e a superfície terrestre. Conforme as características da varredura, o feixe LASER poderá ser refletido por linhas elétricas de alta tensão, vegetação, telhados etc. A classificação da nuvem em pontos no terreno (*terrain*) e nos objetos (*off-terrain*) é denominada filtragem (KOBLER *et al.*, 2007).

Os algoritmos criados para possibilitar a filtragem podem ser classificados de acordo com características operacionais. De acordo com Sithole e Volsselman (2004), as etapas de separação da nuvem de pontos e outros processamentos manuais necessários na produção de um DTM consomem entre 60 e 80% do tempo total necessário ao processamento das observações oriundas do ALS. Estes autores dividiram os algoritmos de filtragem em três grupos:

- Filtro de declividade: o funcionamento deste algoritmo reside no fato de que na diferença de altura entre dois pontos, a probabilidade de um ponto mais alto não pertencer ao terreno aumenta com a diminuição da distância linear em relação à dos pontos próximos, parametrizadas as máximas distâncias e a diferença de altura;
- Filtro de Superfície: um determinado ponto pertence ao terreno se para qualquer ponto pertencente ao terreno, a diferença de altura entre os pontos é menor que a máxima diferença fornecida, que é função da distância entre os pontos;
- Filtro de segmentação: a nuvem de pontos é segmentada e depois classificada, considerando-se que os pontos de um grupamento (*cluster*) não pertencem ao terreno se a altura do segmento estiver acima dos segmentos na vizinhança.

Como os métodos de filtragem são automatizados e parametrizados para situações específicas, os MDT's resultantes podem apresentar falhas em algumas áreas, requerendo a correção manual dos resultados. Áreas íngremes cobertas por florestas são consideradas um problema para a filtragem.

O estado da arte na filtragem de pontos LASER é apresentado por Kobler *et al.* (2007). Estes autores sugerem agrupar os filtros em quatro grupos.

O primeiro grupo de filtros é denominado morfológico, este nome deriva da morfologia matemática. Seu funcionamento consiste em um elemento estrutural que analisa diferenças de alturas na nuvem de pontos em relação a uma distância horizontal. O elemento estrutural pode ser obtido a partir de amostras no terreno ou assumindo um valor de máxima declividade. Este grupo pode trabalhar diretamente



sobre a nuvem de pontos ou sobre a imagem de intensidade. Neste caso, podem ser empregadas técnicas de erosão e de dilatação.

O segundo grupo de filtros trabalha progressivamente, sendo que mais e mais pontos são classificados como pontos no solo (*ground points*). Na primeira etapa os pontos são separados em uma grade regular e em cada célula da grade os pontos com menores altitudes são empregados para gerar uma triangulação. Os pontos classificados como *off ground* são comparados com a triangulação e a análise do ângulo entre os lados dos triângulos são obtidos. A Figura 7 apresenta um ponto B2 classificado como *ground point* quando comparado à triangulação.

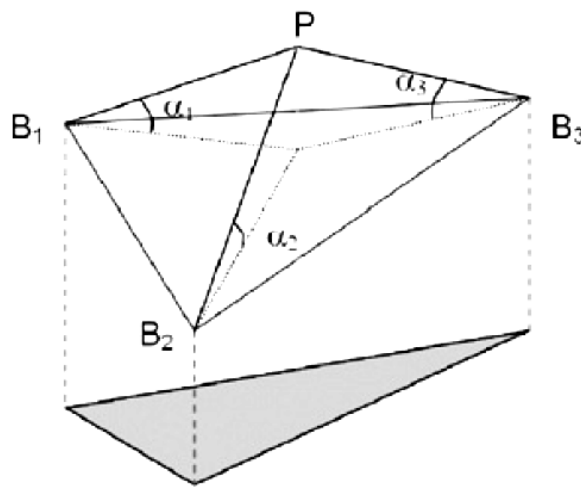


FIGURA 7 - FUNCIONAMENTO DO FILTRO PROGRESSIVO  
FONTE: Kobler *et al.*(2007)

O terceiro grupo de filtros é baseado em uma superfície de referência. Uma primeira superfície é ajustada e usada para determinar os resíduos dos pontos em relação a esta superfície. Se um ponto for classificado corretamente como acima do solo ele terá pouca influência na forma da segunda superfície de referência a ser gerada a partir dos pontos selecionados na primeira etapa. Para determinar a influência de cada ponto, KOBLEK *et al.* (2007) propõem que seja empregado um mecanismo de predição linear, que pode ser obtido pelo emprego de Kriging para a modelagem da superfície. Esta modelagem permite a análise dos componentes no vario grama e desta forma podem ser obtidos pesos baseados nos resíduos calculados pelo vario grama.

O quarto grupo de filtros emprega segmentação de imagens. Neste grupo de filtros Sithole (2005), descreve um método que classifica a segmentação baseada

nas diferenças de altura dos pontos vizinhos. Outra variação deste tipo de filtro seria empregar o crescimento de região associando a variações das alturas para os pontos próximos (NARDINOCCHI *et al.*, 2003). Os filtros baseados em segmentação trabalham sobre grandes entidades, isto é, não consideram *pixels* ou pontos isolados e são desta forma menos influenciados por pontos próximos do solo e que provoquem desta forma erros (ruído) na filtragem (KOBLER *et al.*, 2007).

Além dos grupos apresentados acima, uma nova técnica de filtragem baseada na interpolação repetitiva (REIN), constitui-se em uma ferramenta complementar a outros filtros, especialmente em áreas íngremes e cobertas por florestas. Nestas áreas, outros algoritmos de filtragem têm problemas para distinguir entre os retornos do LASER no solo e na vegetação. A REIN é aplicável depois de uma filtragem inicial e que removeu, não necessariamente, todos os pontos acima do solo. Enquanto os demais filtros trabalham de forma determinística, sempre gerando um erro em situações especiais, com a REIN estes erros não ocorrerão devido ao fato da filtragem estar baseada em um processo aleatório. Kobler *et al.* (2007) desenvolveram a REIN e a utilizaram em uma área íngreme de 2 hectares coberta por floresta mista. Comparando os resultados obtidos com pontos de verificação, os autores determinaram um erro médio de 0,16 metro na determinação de altitudes.

### 2.5.2 Utilização de Dados LASER Scanner na Classificação de Tipologias Florestais

Com relação às formações vegetais naturais, a área de estudo encontra-se no município de União da Vitória no estado do Paraná e os fragmentos da vegetação original fazem parte do Bioma Mata Atlântica, separados em diferentes estágios sucessionais. Para Cullen *et al.* (2003), entende-se por fragmentação o grau de ruptura de uma unidade da paisagem inicialmente contínua. Um dos processos que levou à fragmentação da Floresta foi o cultivo de extensas áreas agrícolas, resultando em fragmentos de diversos tamanhos, formas e distâncias entre eles.

A classificação das formações vegetais utilizadas em sensoriamento remoto, surgidas na década de 1980 têm sido consideradas como fontes primárias de informações acerca da biodiversidade na escala de paisagem, as quais possibilitam

a compreensão da heterogeneidade e dinâmica da paisagem através de mapas temáticos que discriminam a cobertura vegetal de uma área. Recentemente, a identificação e o mapeamento de padrões espaciais das comunidades vegetais têm recebido novos impulsos frente às questões de mudança climática global, ecologia da paisagem e perda da biodiversidade (ZAK; CABIDO, 2002).

Neste contexto, Paiva *et al.* (2007) procederam à classificação da cobertura vegetal na bacia do rio Itapemirim no sul do estado do Espírito Santo empregando imagens digitais, obtidas pelo sensor CCD (*coupled charge device*) embarcado no satélite CBERS 2, associado a informações altimétricas obtidas pela missão topográfica do ônibus espacial (SRTM). Os agrupamentos obtidos pela classificação conjunta dos dados foram cruzados com informações de campo: parâmetros edáficos e fatores climáticos representados pela temperatura, precipitação, déficits hídricos e relação de evapotranspiração real e potencial.

Com relação à classificação das tipologias vegetais por meio da utilização de informações provenientes de sensores remotos, Dlugosz *et al.* (2007) destacam que a utilização de dados de sensoriamento remoto vem possibilitando avaliar e monitorar a preservação de áreas de vegetação natural, além de apresentar grande utilidade na análise do uso e cobertura do solo, gerando informações que fornecem subsídios necessários ao planejamento e administração da ocupação ordenada e racional do meio físico. Com os produtos resultantes dos sensores ópticos, as linhas de pesquisa que avaliam estas imagens, voltadas para o campo florestal, têm abordado operações de segmentação, baseada em algoritmos de crescimento de regiões baseado em morfologia matemática. Por meio do processo de segmentação é possível obter-se, de maneira rápida, a discriminação e delimitação de fragmentos. Adicionalmente, os polígonos gerados pela segmentação, podem ser usados como amostras de treinamento em futuras classificações digitais e, ainda, em classificações visuais por meio de simples atribuição de classes aos polígonos ou ainda por associação dos polígonos através de trabalhos a campo empregando técnicas de inventário florestal.

Analisando o estado da arte em classificação da nuvem de pontos LASER em classes de uso e ocupação do solo, verifica-se, a nível nacional, que esta área de pesquisa está pouco explorada, pois normalmente as informações da varredura gerada pelo ALS são combinadas com outras fontes de dados espectrais. A seguir são apresentados alguns trabalhos nesta área.

Kim *et al.* (2009) afirmam que a identificação das espécies arbóreas é importante para uma variedade de atividades de gerenciamento e monitoramento de recursos naturais incluindo caracterização de matas ciliares, risco de incêndio florestal, monitoramento de biodiversidade e *habitat*. Uma identificação das espécies arbóreas pode ser conduzida por um trabalho de inventário no campo ou através da interpretação de fotografias aéreas em escala grande. Contudo, estes métodos são caros, consumindo tempo e não aplicáveis para áreas grandes ou isoladas. Com o advento do sensoriamento remoto orbital foram desenvolvidos vários esforços no sentido de suprimir esta lacuna; contudo, a utilização deste tipo de dados tem limitações para distinguir espécies florestais, devido ao grande número de bandas espectrais, para o caso, de imagens hiperespectrais. Os sistemas LIDAR oferecem vantagens em relação a outras técnicas de sensoriamento remoto, principalmente, sua habilidade de capturar medidas tridimensionais sobre grandes áreas. O LASER *scanner* é normalmente utilizado para a obtenção de medidas com a finalidade de geração automatizada do modelo digital do terreno ou um modelo digital de copas. Esta tecnologia pode suprir atividades de monitoramento e gerenciamento florestal com acurácia e eficiência.

Até bem pouco tempo, as pesquisas envolvendo as mensurações LIDAR estavam focadas na estimação da estrutura florestal, citando como exemplo a altura, volume e biomassa de áreas cobertas por espécies arbóreas. Contudo, a capacidade de registrar a intensidade, que é a medida da porção de energia refletida pelos objetos no instante da varredura LASER, habilita esta tecnologia na discriminação da cobertura vegetal.

Morsdorf *et al.* (2005) usaram a extração do índice foliar de vegetação (IFV) empregando recobrimento LIDAR a baixas alturas (850 m) e com alta densidade de pontos (10 pontos/m<sup>2</sup>). Tomando a diferença entre o primeiro e último retorno como parâmetro discreto de análise da cobertura foliar, foi possível a análise e determinação do diâmetro médio das árvores empregando as fórmulas do inventário florestal.

Blair *et al.* (1999) descrevem a utilização de um ALS desenvolvido pela NASA especificamente para a realização de mapeamento topográfico, hidrológico e de vegetação com acuracidade e eficiência. Nos testes realizados os autores verificaram a capacidade para a determinação da altura da cobertura vegetal e conseqüentemente a caracterização vertical das formações florestais.

A discriminação de espécies florestais foi testada por Moffiet *et al.* (2005) que, realizando a varredura LIDAR sobre uma área de floresta subtropical no sudeste da Austrália, concluíram que a proporção do sinal de retorno do LASER e o índice de permeabilidade da vegetação são parâmetros que podem ser utilizados para a separação de espécies florestais.

A determinação de biomassa florestal, que é definida como sendo a quantidade de volume seco de um grupo de árvores, pode ser obtida através do processamento da varredura realizada pelo ALS. Os autores Bortolot e Wynne (2005), testaram esta possibilidade em florestas naturais e plantadas localizadas no estado da Virginia (USA) e com isso conseguiram determinar a biomassa florestal para uma área de teste com valores bem próximos aos verificados em campo.

Pfeifer *et al.* (2004), ao analisarem a influência da cobertura vegetal nas mensurações obtidas pelo LIDAR, verificaram que as formações vegetais de altura baixa (arbustos e sebes como exemplos) causam deslocamentos verticais da ordem de 5 a 17 centímetros nas mensurações. Acrescenta-se que este tipo de cobertura vegetal não é detectado adequadamente pelas técnicas de classificação usualmente empregadas. Estes autores propuseram estratégias para a correção deste deslocamento vertical, sem fontes adicionais de dados relativos ao tipo da cobertura vegetal. Para tanto, propuseram a classificação da nuvem de pontos baseada na textura, que são repetições regulares de um fenômeno espacial. Os valores de textura foram computados para a nuvem de pontos LASER e relacionados com o tipo de cobertura vegetal tornando possível a correção das mensurações através da correlação textura e deslocamento vertical. A hipótese empregada assume que quando diferentes classes de cobertura vegetal têm textura similar, elas também têm influência similar sobre as medidas LASER e desta forma requerem correções similares. Por exemplo, arbustos têm valor de textura maior (mais grosseira) do que grama com textura menor (mais fina) e necessitarão de uma correção vertical proporcionalmente maior.

Orka *et al.* (2009) realizaram a classificação individualizada de árvores de acordo com a espécie florestal utilizando observações geradas por um ALS marca OPTTECH modelo ALTM 3100. A área de estudo localiza-se em uma reserva natural na região sudeste da Noruega. Neste local foram implantadas vinte parcelas e realizado o inventário florestal georreferenciado que possibilitou a classificação *in locu* das árvores, de acordo com suas espécies, nas parcelas. A partir das

observações geradas pelo inventário foram escolhidas amostras na nuvem de pontos LIDAR. Estas amostras possibilitaram o ajuste das funções de discriminação linear, obtidas através da técnica de regressão linear. Os cofatores de classificação escolhidos para o ajuste das funções: média e desvio padrão da intensidade além da altura da cobertura vegetal, obtida a partir do Modelo Digital de Superfícies Normalizado. Os autores classificaram a nuvem de pontos e, comparando os resultados da classificação com o inventário de campo, obtiveram índices de acerto que variaram de 57,7 a 93,3%.

Um estudo similar pode ser verificado em Orka *et al.* (2007).

Empregando funções de discriminação do tipo linear e cofatores relacionados à intensidade média nas partes superior, média e inferior das copas das árvores, Kim *et al.* (2009) realizaram a classificação das espécies arbóreas no Parque Arboreto de Washington. Para tanto, foram utilizadas mensurações obtidas por duas varreduras LASER, uma no inverno e outra no verão, possibilitando desta forma a obtenção do modelo digital do terreno e também a modelagem tridimensional das árvores isoladas. Os pontos LIDAR de cada uma das árvores foram identificados a partir do posicionamento das árvores obtido a campo. A intensidade média de diferentes espécies foi comparada, considerando além das características reflexivas da vegetação, a presença ou ausência de folhas e seu arranjo nas copas das árvores. A classificação da nuvem de pontos foi conduzida empregando-se funções de discriminação do tipo linear e o acerto, quando comparado com as mensurações de campo, variou de 54,2 a 91,0%. Este estudo destaca que a intensidade média pode ser utilizada efetivamente na discriminação de uma ampla variedade de espécies florestais naturais.

## 2.6 ESTATÍSTICA MULTIVARIADA

A Estatística Multivariada consiste em um conjunto de métodos estatísticos utilizados em situações nas quais várias variáveis são medidas simultaneamente, em cada elemento amostral. A Estatística ou Métodos Multivariados se divide em dois grupos: o primeiro consistindo em técnicas exploratórias de sintetização da estrutura de variabilidade dos dados, e o segundo consistindo em técnicas de

inferência estatística. Em linhas gerais os métodos de Estatística Multivariada são utilizados com o intuito de simplificar ou facilitar a interpretação do fenômeno que está sendo estudado através da construção de índices ou variáveis alternativas que sintetizem a informação original dos dados (MINGOTI, 2005).

Segundo Johnson (1998), os métodos multivariados são úteis para auxiliar na análise de grande quantidade de informações provenientes de um grupo de variáveis obtidas a partir de certo número de experimentos. A importância e uso dos métodos multivariados aumentam com o número de variáveis medidas, bem como, com o aumento do número de experimentos.

Os dados multivariados são procedentes da coleta de  $n$  observações com  $p \geq 1$  variáveis ou características medidas. A representação destes dados amostrais é feita através de uma matriz correspondente às  $n$  observações do vetor aleatório  $X$   $p$ -dimensional:

$${}_n X_p = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad (4)$$

O vetor aleatório  $\underline{X} = [X_1 \ X_2 \ \dots \ X_p]$  de dimensão  $p$  foi observado  $n$  vezes dando origem à matriz de dados  $X$  de ordem  $n \times p$ .

Paixão (2006) cita exemplos de técnicas multivariadas, entre elas, Análise de Componentes Principais, Análise Fatorial e Análise de Correlação Canônica. Também cita as técnicas que trabalham com o reconhecimento e classificação de padrões: Análise Discriminante e Regressão Logística, além de análise de agrupamentos de entidades vetoriais.

### 2.6.1 Reconhecimento de Padrões

O reconhecimento de padrões é um conjunto de técnicas capaz de classificar, associar e agrupar elementos através de seus atributos.

As principais abordagens de Reconhecimento de Padrões são: a abordagem estatística (assume que as características das classes se regem por determinados

modelos probabilísticos) e a abordagem sintática ou estrutural (procura descrever a estrutura dos padrões usando inter-relações de características descritivas básicas denominadas primitivas). Mais recentemente surgiu a tecnologia de Redes Neurais e também métodos de Programação Matemática.

O problema clássico em Reconhecimento de Padrões pode ser resumido como: “Dado um vetor de medidas ( $m_i$ ), obter um método de inversão de mapeamento nas relações  $g$  (função geradora) e  $m$  (função de mapeamento), de forma que seja possível identificar a classe geradora de medidas” (PAIXÃO, 2006). A figura 8 apresenta o princípio de funcionamento do Reconhecimento de Padrões.

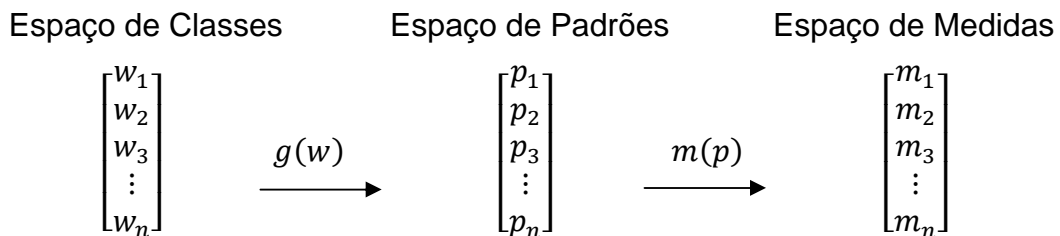


FIGURA 8 - PROBLEMA CLÁSSICO DE RECONHECIMENTO DE PADRÕES  
FONTE: Adaptado de Paixão (2006).

O reconhecimento de padrões extrai as características de um elemento e uma próxima etapa caracteriza-o. A característica é um vetor de medidas do elemento que se deseja classificar.

Deve ainda ser considerado no escopo do reconhecimento de padrões o arranjo dos grupos de dados a serem analisados. Este fato torna-se mais relevante com a Regressão Logística, pois está relacionado a estimação dos parâmetros incógnitos do modelo a ser determinado.

A configuração dos dados refere-se a separação entre eles. Albert e Anderson (1984) sugere três situações para a separação dos grupos. A Figura 9 apresenta uma possível configuração empregando 3 grupos.



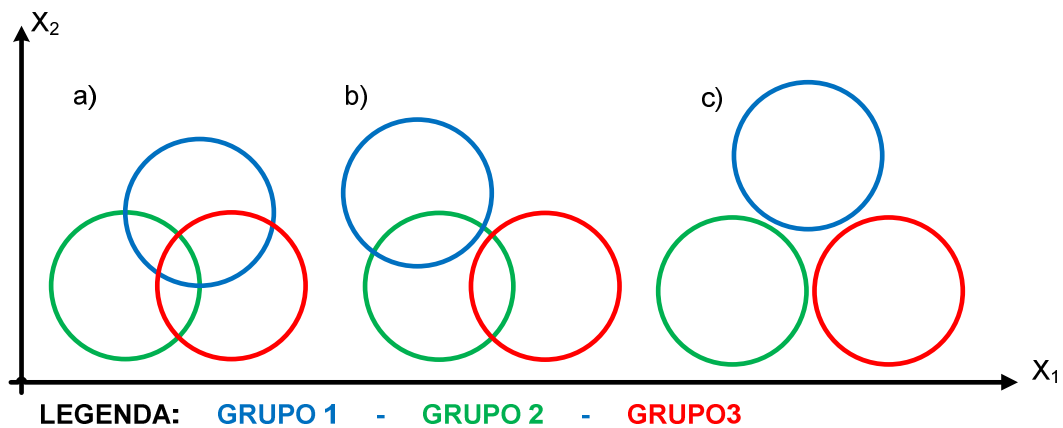


FIGURA 9 - CONFIGURAÇÃO DE GRUPOS  
 FONTE: Adaptado de Guimarães, 2006.

A Figura 9(a) apresenta três grupos com sobreposição. A situação apresentada em 9(b) mostra uma situação de sobreposição parcial e a 9(c) mostra grupos completamente separados, ou sem sobreposição entre as variáveis.

### 2.6.2 Análise Discriminante

A Análise Discriminante é uma técnica estatística que pode ser empregada para reconhecer e classificar elementos de uma amostra ou população. Para a sua aplicação é necessário que os grupos, para os quais cada elemento da amostra possa ser classificado, sejam conhecidos de acordo com suas características gerais. Este conhecimento permite a elaboração de uma função matemática chamada de regra de classificação ou discriminação (MINGOTI, 2005).

Suponha que se tenha  $n_1$  elementos amostrais procedentes, com probabilidade 1, da população A e  $n_2$  elementos amostrais procedentes, com probabilidade 2, da população B, e que para cada um dos  $n_1 + n_2 = n$  elementos amostrais tenham sido medidas  $p$  - variáveis aleatórias (características). A análise estatística do comportamento das  $p$  - variáveis medidas permite identificar o perfil geral de cada grupo. Deste modo, se houver um novo elemento amostral, não pertencente a nenhuma das duas amostras anteriores, e cuja origem é incerta, seria

possível compará-lo de algum modo com o perfil geral dos grupos A e B e classificá-los como pertencente ao grupo cujo perfil geral foi mais semelhante ao dele.

A análise discriminante teve suas primeiras aplicações na área de Botânica, sendo empregada na discriminação de grupos de plantas através dos parâmetros tamanho e tipo de folhas, sendo então possível posteriormente, classificar as novas espécies encontradas. Em 1936, Fisher foi o responsável pelo desenvolvimento da análise para dois ou mais grupos.

A seguir são apresentados os conceitos e formulação das respectivas funções.

#### 2.6.2.1 Função Discriminante Linear de Fisher para duas populações

A Função Discriminante Linear de Fisher obtém novas variáveis através de combinações lineares das variáveis originais. As variáveis discriminantes são criadas de forma que contenham todas as informações a respeito do conjunto das variáveis originais (JOHNSON, 1998).

Uma vantagem da Função Discriminante Linear de Fisher ou Função Discriminante Canônica é permitir a interpretação da distância entre as populações e também possibilitar a redução do espaço dimensional.

A idéia inicial de Fisher foi transformar as observações multivariadas  $\underline{X}$  nas observações univariadas  $Y$  tal que os  $Y$ 's as populações  $\pi_1$  e  $\pi_2$  fossem separadas tanto quanto possível. Isto é feito fazendo  $Y$  como combinação linear de  $X$ .

A seguir é apresentado o desenvolvimento do método de Fisher para a classificação em duas populações. O desenvolvimento está de acordo com a explanação dada por Paixão (2006).

Sejam:

- $\pi_1$  e  $\pi_2$ , duas populações cujos elementos consistem nos vetores formados por  $p$  variáveis aleatórias representadas pelo vetor aleatório  $\underline{X} = [x_1 \ x_2 \ \dots \ x_p]$ ;
- $\mu_{1Y}$  é a média dos  $Y$ 's obtidos dos  $X$ 's pertencentes à população  $\pi_1$  e  $\mu_{2Y}$  que é a média dos  $Y$ 's obtidos dos  $X$ 's pertencentes à população  $\pi_2$ .

Fisher selecionou a combinação linear que maximiza a distância quadrática entre  $\mu_{1Y}$  e  $\mu_{2Y}$  com relação a variabilidade dos  $Y$ 's.

Sejam ainda:

$$\underline{\mu}_1 = E(\underline{X}|\pi_1): \text{valor esperado de uma observação multivariada de } \pi_1; \quad (5)$$

$$\underline{\mu}_2 = E(\underline{X}|\pi_2): \text{valor esperado de uma observação multivariada de } \pi_2; \quad (6)$$

E ainda a matriz de covariâncias:

$$\Sigma = E[(\underline{X} - \underline{\mu}_i)(\underline{X} - \underline{\mu}_i)'] \quad i = 1, 2 \quad (7)$$

Neste caso, considerando a mesma matriz de covariâncias para ambas as populações ( $\Sigma_1 = \Sigma_2$ ) tem-se a combinação linear:

$$Y = \underline{c}'\underline{X} \quad (8)$$

que representa o vetor  $\underline{X}$  em uma única dimensão.

Substituindo-se a equação (7) nas equações (5) e (6), obtêm-se:

$$\mu_{1Y} = E(Y|\pi_1) = E(\underline{c}'\underline{X}|\pi_1) = \underline{c}'E(\underline{X}|\pi_1) = \underline{c}'\underline{\mu}_1 \quad (9)$$

e

$$\mu_{2Y} = E(Y|\pi_2) = E(\underline{c}'\underline{X}|\pi_2) = \underline{c}'E(\underline{X}|\pi_2) = \underline{c}'\underline{\mu}_2 \quad (10)$$

Considerando, ainda, a variância:

$$\sigma_Y^2 = V(Y) = V(\underline{c}'\underline{X}) = \underline{c}'V(\underline{X})\underline{c} = \underline{c}'\Sigma\underline{c} \quad (11)$$

têm-se que a melhor combinação linear vem da razão entre o quadrado da distância entre as médias e a variância de  $Y$ , ou seja,  $\underline{c}$  é tal que maximiza a razão:

$$\frac{(\mu_{1Y} - \mu_{2Y})^2}{\sigma_Y^2} = \frac{(\underline{c}'\underline{\mu}_1 - \underline{c}'\underline{\mu}_2)^2}{\underline{c}'\underline{\Sigma}\underline{c}} \quad (12)$$

Fazendo

$$\underline{\delta} = \underline{\mu}_1 - \underline{\mu}_2 \quad (13)$$

$$V(\underline{X}) = \underline{\Sigma} \quad (14)$$

Resulta na equação (15)

$$\frac{(\underline{c}'\underline{\delta})^2}{\underline{c}'\underline{\Sigma}\underline{c}} \quad (15)$$

A equação (15) será maximizada por

$$\underline{c} = k\underline{\Sigma}^{-1} \quad (16)$$

para um  $k \neq 0$  qualquer e

$$\underline{\delta} = k\underline{\Sigma}^{-1}(\underline{\mu}_1 - \underline{\mu}_2) \quad (17)$$

Fazendo  $k = 1$ , já que pode ser qualquer valor e para simplificar, têm-se que:

$$\underline{c} = \underline{\Sigma}^{-1}(\underline{\mu}_1 - \underline{\mu}_2) \quad (18)$$

e

$$Y = \underline{c}'\underline{X} = (\underline{\mu}_1 - \underline{\mu}_2)'\underline{\Sigma}^{-1}\underline{X} \quad (19)$$

A equação (19) é a Função Discriminante Linear de Fisher (FDL).

Considere-se agora o processo de classificação de uma nova observação  $\underline{X}_0$ . Qual seria o resultado da função? A nova amostra seria alocada na população  $\pi_1$  ou  $\pi_2$ ?

Partindo da equação (19), obtêm-se:

$$Y_0 = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{X}_0 \quad (20)$$

Seja  $m$  o ponto médio entre as duas populações univariadas:

$$m = \frac{1}{2}(\mu_{1Y} + \mu_{2Y}) = \frac{1}{2}(\underline{c}'\underline{\mu}_1 + \underline{c}'\underline{\mu}_2) = \frac{1}{2}\underline{c}'(\underline{\mu}_1 + \underline{\mu}_2) = \frac{1}{2}\underline{c}'(\underline{\mu}_1 - \underline{\mu}_2) \Sigma^{-1}(\underline{\mu}_1 + \underline{\mu}_2) \quad (21)$$

$$\text{e ainda: } \underline{c}' = (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1}$$

É possível mostrar que:

$$E(Y_0 | \pi_1) - m = E((\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} (\underline{X}_0 | \pi_1)) - \frac{1}{2}(\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) = \\ (\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} \underline{\mu}_1 - \frac{1}{2}(\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 + \underline{\mu}_2) = \frac{1}{2}(\underline{\mu}_1 - \underline{\mu}_2)' \Sigma^{-1} (\underline{\mu}_1 - \underline{\mu}_2) \geq 0 \quad (22)$$

ou seja,

$$E(Y_0 | \pi_1) - m \geq 0 \quad (23)$$

e ainda

$$E(Y_0|\pi_2) - m < 0 \quad (24)$$

A partir do exposto nas equações (23) e (24),  $\underline{X}_0 \in \pi_1$  desde que  $\underline{Y}_0$  seja maior ou igual a  $m$  e, de maneira análoga  $\underline{X}_0 \in \pi_2$  desde que  $\underline{Y}_0$  seja menor que  $m$ .

A regra de classificação resulta em:

$\underline{X}_0$  pertencente a população  $\pi_1$  se  $Y_0 \geq m$

$\underline{X}_0$  pertencente a população  $\pi_2$  se  $Y_0 < m$

Na realidade, os parâmetros populacionais não são conhecidos e a função discriminante linear deverá ser estimada. Desta forma:

$$\hat{Y} = \hat{c}'X = (\bar{\underline{X}}_1 - \bar{\underline{X}}_2)'S_p^{-1}\underline{X} \quad (25)$$

o ponto médio  $m$  pode ser estimado por:

$$m = \frac{1}{2}(\bar{\underline{X}}_1 - \bar{\underline{X}}_2)'S_p^{-1}(\bar{\underline{X}}_1 + \bar{\underline{X}}_2) = \frac{\bar{Y}_1 + \bar{Y}_2}{2} \quad (26)$$

Sendo que:

$$\bar{\underline{X}}_1 = \frac{1}{n_1} \sum_{i=1}^{n_1} X_{i1} \quad (27)$$

A equação (27) é o estimador de  $\mu_1$ .

$$\bar{\underline{X}}_2 = \frac{1}{n_2} \sum_{i=1}^{n_2} X_{i2} \quad (28)$$

A equação (28) estima o valor de  $\mu_2$ .

O estimador de  $\Sigma$  é calculado por:

$$S_p = \frac{(n_1-1)S_1 + (n_2-1)S_2}{n_1 + n_2 - 2} \quad (29)$$

Onde:

$$S_1 = \frac{1}{n_1-1} \sum_{i=1}^{n_1} (\underline{X}_{i1} - \bar{\underline{X}}_1) (\underline{X}_{i1} - \bar{\underline{X}}_1)' \quad (30)$$

e

$$S_2 = \frac{1}{n_2-1} \sum_{i=1}^{n_2} (\underline{X}_{i2} - \bar{\underline{X}}_2) (\underline{X}_{i2} - \bar{\underline{X}}_2)' \quad (31)$$

A regra de alocação para uma observação considerando agora a função discriminante aproximada resultará em:

$\underline{X}_0$  será pertencente a população  $\pi_1$  se  $\hat{Y}_0 - \hat{m} \geq 0$  ou  $\hat{Y}_0 \geq \hat{m}$

$\underline{X}_0$  será pertencente a população  $\pi_2$  se  $\hat{Y}_0 - \hat{m} < 0$  ou  $\hat{Y}_0 < \hat{m}$

#### 2.6.2.2 Função Discriminante Linear de Fisher para diversas populações

O método de Fisher não necessita que as populações apresentem distribuição normal de probabilidades, contudo, as covariâncias populacionais são admitidas como iguais:

$$\Sigma_1 = \Sigma_2 = \dots \Sigma_g = \Sigma$$

O vetor médio das diversas populações resulta em:

$$\underline{\bar{\mu}} = \frac{1}{g} \sum_{i=1}^g \underline{\mu}_i \quad (32)$$

Considere-se ainda a matriz soma de produtos cruzados entre grupos populacionais ( $B_0$ ):

$$B_0 = \sum_{i=1}^g (\underline{\mu}_i - \underline{\bar{\mu}})(\underline{\mu}_i - \underline{\bar{\mu}})' \quad (33)$$

A combinação linear  $Y = \underline{c}'\underline{X}$  tem esperança matemática por:

$$E(Y) = \underline{c}'E(\underline{X}|\pi_i) = \underline{c}'\underline{\mu}_i \quad (34)$$

sendo que  $i = 1, 2, \dots, g$  (populações/grupos).

A variância pode ser calculada por:

$$\sigma_Y^2 = V(Y) = V(\underline{c}'\underline{X}) = \underline{c}'V(\underline{X})\underline{c} = \underline{c}'\underline{\Sigma}\underline{c} \quad (35)$$

Desta forma o valor esperado  $\mu_{iY} = \underline{c}'\underline{\mu}_i$  se altera quando a população da qual  $\underline{X}$  é seleccionado é outra (PAIXÃO, 2006).

A média global poderá ser determinada por:

$$\bar{\mu}_Y = \frac{1}{g} \sum_{i=1}^g \mu_{iY} = \frac{1}{g} \sum_{i=1}^g \underline{c}'\underline{\mu}_i = \underline{c}'\underline{\bar{\mu}} \quad (36)$$

O que resulta na razão entre a soma dos quadrados das distâncias das populações para a média global de Y e a variância de Y:

$$\frac{\sum_{i=1}^g (\mu_{iY} - \bar{\mu}_Y)^2}{\sigma_Y^2} = \frac{\sum_{i=1}^g (\underline{c}'\underline{\mu}_i - \underline{c}'\underline{\bar{\mu}})^2}{\underline{c}'\underline{\Sigma}\underline{c}} = \frac{\sum_{i=1}^g (\underline{c}'(\underline{\mu}_i - \underline{\bar{\mu}})(\underline{\mu}_i - \underline{\bar{\mu}})'\underline{c})}{\underline{c}'\underline{\Sigma}\underline{c}}$$



$$\frac{\underline{c}' [\sum_{i=1}^g (\underline{\mu}_i - \underline{\bar{\mu}})(\underline{\mu}_i - \underline{\bar{\mu}})'] \underline{c}}{\underline{c}' \underline{\Sigma} \underline{c}} = \frac{\underline{c}' B_0 \underline{c}}{\underline{c}' \underline{\Sigma} \underline{c}} \quad (37)$$

A equação (37) mede a variabilidade entre os grupos de valores para  $Y$ , relativos à variabilidade comum entre os grupos.

Para resolver o problema deve-se seleccionar o  $\underline{c}$  que maximiza a equação (37).

Fazendo a padronização:

$$\underline{c}' \underline{\Sigma} \underline{c} = 1$$

e ainda  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_s > 0$  são os  $s \leq \min(g-1, p)$  autovalores não nulos de  $\Sigma^{-1} B_0$  e  $\underline{e}_1, \underline{e}_2, \dots, \underline{e}_s$  os respectivos autovetores de tal maneira que  $\underline{e}' \underline{\Sigma} \underline{e} = 1$ .

O coeficiente  $\underline{c}$  que maximiza a razão mostrada na equação (37) é dada por  $\underline{c}_1 = \underline{e}_1$ .

Denomina-se o primeiro discriminante a combinação linear:

$$\underline{c}_1' \underline{X} = \underline{e}_1' \underline{X}$$

e segundo discriminante

$$\underline{c}_2' \underline{X} = \underline{e}_2' \underline{X}$$

Sendo que  $\underline{c}_2 = \underline{e}_2$  maximiza a  $\frac{\underline{c}' B_0 \underline{c}}{\underline{c}' \underline{\Sigma} \underline{c}}$ , desde que a condição  $COV(\underline{c}_1' \underline{X}, \underline{c}_2' \underline{X}) = 0$  e assim as várias combinações lineares referentes a cada população podem ser obtidas (PAIXÃO, 2006).

Normalmente a variância e média populacional não são conhecidas e suas estimativas podem ser obtidas através de amostras de tamanhos  $n_{em}$  das  $g$  populações.

Desta forma o estimador de  $\mu_i$  pode ser calculado por:

$$\bar{X}_i = \frac{1}{n} \sum_{j=1}^{n_i} X_{ij} \quad (38)$$

O estimador da média global resultará em:

$$\bar{X} = \frac{\sum_{i=1}^g n_i \bar{X}_i}{\sum_{i=1}^g n_i} = \frac{\sum_{i=1}^g \sum_{j=1}^{n_i} X_{ij}}{\sum_{i=1}^g n_i} \quad (39)$$

E a soma de produtos cruzados entre grupos populacionais é:

$$\hat{B}_0 = \sum_{i=1}^g (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})' \quad (40)$$

A matriz  $\Sigma$  pode ser calculada através de:

$$W = \sum_{i=1}^g \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)' \quad (41)$$

Logo,  $S_p$  resulta em:

$$S_p = \frac{W}{n_1 + n_2 + \dots + n_g - g} = \frac{(n_1 - 1)S_1 + (n_2 - 1)S_2 + \dots + (n_g - 1)S_g}{n_1 + n_2 + \dots + n_g - g} \quad (42)$$

O vetor  $\underline{\hat{c}}$  que maximiza a razão:

$$\frac{\underline{\hat{c}}' \hat{B}_0 \underline{\hat{c}}}{\underline{\hat{c}}' S_p \underline{\hat{c}}} \quad (43)$$

É o mesmo que maximiza a equação:

$$\frac{\underline{\hat{c}}' \hat{B}_0 \underline{\hat{c}}}{\underline{\hat{c}}' W \underline{\hat{c}}} \quad (44)$$

E o vetor  $\underline{\hat{c}}$  é obtido da matriz  $W^{-1} \hat{B}_0$ .

Se  $W^{-1} \hat{B}_0 \underline{\hat{e}} = \hat{\lambda} \underline{\hat{e}}$  então

$$S_p^{-1} \hat{B}_0 \underline{\hat{e}} = \hat{\lambda} (n_1 + n_2 + \dots + n_g - g) \underline{\hat{e}} \quad (45)$$

E ainda

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \dots \lambda_s > 0$$

Os autovalores e autovetores da  $W^{-1} \hat{B}_0$ , sendo  $s \leq \min(g-1, p)$  e  $\underline{\hat{e}}_i$  o autovetor normalizado de forma que:

$$\underline{\hat{e}}_i' S_p \underline{\hat{e}}_i = 1 \quad (46)$$

O vetor de coeficientes que maximiza a razão  $\frac{\underline{\hat{c}}' \hat{B}_0 \underline{\hat{c}}}{\underline{\hat{c}}' W \underline{\hat{c}}}$  será  $\underline{\hat{c}}_1 = \underline{\hat{e}}_1$  sendo este

o primeiro discriminante amostral:

$$\underline{\hat{c}}_1' X \quad (47)$$

### 2.6.2.3 Função Discriminante Quadrática

Supondo as matrizes de covariância de duas populações,  $\Sigma_1$  para  $\underline{x} \in \Pi_1$  e  $\Sigma_2$  para  $\underline{x} \in \Pi_2$  com  $\Sigma_1 \neq \Sigma_2$ . Esta última condição torna as regras de reconhecimento de padrões mais complicadas. Sejam então  $\underline{x} \sim N_p(\underline{\mu}_i, \Sigma_i)$ , com  $i = 1, 2$  e  $\underline{\mu}_1 \neq \underline{\mu}_2$  e  $\Sigma_1 \neq \Sigma_2$ . A probabilidade total de reconhecimento errado e o custo esperado de reconhecimento errado dependem da razão de densidades dada por:

$$\frac{f_1(\underline{x})}{f_2(\underline{x})} \quad (48)$$

que é equivalente ao logaritmo das razões de densidades:

$$\ln \left[ \frac{f_1(\underline{x})}{f_2(\underline{x})} \right] = \ln[f_1(\underline{x})] - \ln[f_2(\underline{x})] \quad (49)$$

A partir do exposto, pode-se realizar o desenvolvimento da expressão da função discriminante quadrática para várias populações (PAIXÃO, 2006). Sejam os vetores de variáveis aleatórias  $\underline{X}$  de populações com distribuição normal e multivariada  $(\underline{\mu}_i, \Sigma_i)$ :

$$f_i(\underline{X}) = \frac{1}{(2\pi)^{p/2} |\Sigma_i|} \exp \left[ -\frac{1}{2} (\underline{X} - \underline{\mu}_i)' \Sigma_i^{-1} (\underline{X} - \underline{\mu}_i) \right] \quad (50)$$

Sendo:

$i = 1, 2, \dots, g$

$\underline{\mu}_i$  - vetor das médias da população  $\Pi_i$

$\Sigma_i$  - matriz de covariâncias do vetor  $\underline{x} \in \Pi_i$ .

A regra de classificação será alocar  $\underline{X}$  em  $\pi_k$ , se o score é o maior, ou seja:

$$\ln p_k f_k(\underline{X}) = \ln p_k - \left(\frac{p}{2}\right) \ln(2\pi) - \frac{1}{2} \ln|\Sigma_k| - \frac{1}{2} (\underline{X} - \underline{\mu}_k)' \Sigma_k^{-1} (\underline{X} - \underline{\mu}_k) = \max (\ln p_i f_i) \quad (51)$$

A constante  $\frac{p}{2} \ln(2\pi)$  é a mesma para todas as populações e portanto, pode ser retirada da expressão. O escore (função de discriminação quadrática) resulta então em:

$$d_i^Q(\underline{X}) = \frac{1}{2} \ln|\Sigma_i| - \frac{1}{2} (\underline{X} - \underline{\mu}_i)' \Sigma_i^{-1} (\underline{X} - \underline{\mu}_i) + \ln p_i \quad (52)$$

A regra para classificação de  $\underline{X}$  em  $\pi_k$  se:

$$d_k^Q(\underline{X}) = \max d_i^Q(\underline{X}) \quad (53)$$

Como os parâmetros populacionais não são conhecidos, empregam-se respectivamente os seus estimadores e a função de classificação quadrática resulta em:

$$D_i = \frac{1}{2} \ln|S_i| - \frac{1}{2} \underline{X}' S_i^{-1} \underline{X} + \bar{\underline{X}}_i' S_i^{-1} \underline{X} - \frac{1}{2} \bar{\underline{X}}_i' S_i^{-1} \bar{\underline{X}}_i + \ln(P_i) \quad (54)$$

Sendo:

$D_i$  - escore discriminante quadrático de classificação

$\bar{\underline{X}}_i$  - vetor médio do grupo  $i$

$S_i$  - é o estimador da matriz de covariância populacional do grupo  $i$

$P_i$  - probabilidade de um indivíduo ser classificado no grupo  $i$ .

$i = 1, 2, \dots, g$  (grupos/populações).

#### 2.6.2.4 Função de Discriminação Logística

A regressão logística foi desenvolvida no século XIX para descrever o

crescimento das populações e de reações químicas em autocatálise. O modelo logístico conhecido na época era razoável para se estudar o crescimento de populações de países jovens. Alphonse Quetelet, um astrônomo belga, previa a extrapolação do crescimento exponencial que conduz a valores impossíveis para o parâmetro populacional.

Pierre François Verhulst em 1838 estudando os ajuste de curvas propostos por Quetelet, publicou três artigos sobre esta função, passando a denominá-la função logística.

Em 1920, Raymond Pearl e Lowell Reed estudando o crescimento populacional dos Estados Unidos, empregam a função logística em seu trabalho. A idéia básica do desenvolvimento logístico é simples e efetiva, sendo empregada atualmente para modelar o crescimento populacional e na introdução de novos produtos e tecnologias no mercado, entre eles, telefones celulares e aplicações nas áreas de saúde e finanças.

Antes de prosseguir com os conceitos relacionadas a Regressão Logística é interessante definir probabilidade condicionada.

De acordo com Christensen (1997) a probabilidade condicionada é um conceito importante para o entendimento em Regressão Logística e Análise do Modelo Log-linear. Seja  $E$  um experimento aleatório e  $S$  o espaço amostral associado ao experimento formado por " $n$ " resultados igualmente prováveis. Seja  $A \subseteq S$  um evento com " $m$ " elementos. A probabilidade de  $A$ , anotada por  $P(A)$  é definida como sendo:

$$P(A) = m / n \quad (55)$$

Considere-se agora um evento  $B$ . Se  $Pr(B) > 0$ , a probabilidade condicional de  $A$  desde que o evento  $B$  tenha ocorrido é

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (56)$$

e

$$P(A \cap B) = P(A).P(B) \quad (57)$$

Na equação (57) o termo:  $P(A \cap B)$  calcula a probabilidade de ocorrência do evento A e B simultaneamente.

Ao calcular  $P(B/A)$  determina-se a probabilidade de ocorrência do evento B em relação ao espaço amostral reduzido de A.

Sabendo-se que B ocorre e não influencia a ocorrência de A, então A é independente de B. Especificamente, A é independente de B se

$$P(A|B) = P(A) \quad (58)$$

Alguns exemplos de probabilidade condicional podem ser vistos em Christensen (1997).

#### 2.6.2.4.1 Modelo de Regressão Logístico Simples

Os métodos de regressão tornaram-se uma componente integral da análise de dados, descrevendo a relação entre uma variável resposta e uma ou mais variáveis explanatórias. É importante entender que o objetivo de uma análise empregando a regressão é a mesma que a de qualquer técnica de construção de modelos usados em Estatística: encontrar o melhor ajuste, o modelo que descreve a relação entre uma variável resposta e um conjunto de variáveis independentes (cofatores). O exemplo mais usual deste ajuste é a regressão linear onde a variável resposta é considerada contínua (HOSMER;LEMESHOW, 1989).

De acordo com Hosmer e Lemeshow (1989) as características da regressão logística são:

- A média condicionada da função de regressão é formulada para o intervalo de zero a um;
- A distribuição binomial, descreve a distribuição dos erros para o modelo;
- Os princípios que guiam uma regressão linear também podem ser empregados em regressão logística.

A forma usual do modelo de regressão logística é:

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x}}{1 + e^{\beta_0 + \beta_1 x}} \quad (59)$$

Sendo que  $\pi(x) \in [0,1]$  é a conhecida função sigmóide que varia de zero a um.

Uma transformação de  $\pi(x)$  que é importante para o estudo da regressão logística é a transformação logit:

$$g(x) = \ln \left[ \frac{\pi(x)}{1 - \pi(x)} \right] = \ln(\pi(x)) - \ln(1 - \pi(x)) \quad (60)$$

A importância da transformação mostrada na equação (60) é que muitas das propriedades desejáveis da regressão linear estão presentes:

- $g(x)$  é linear em sua parametrização;
- é contínua e
- pode ter um intervalo de solução variando de  $-\infty$  a  $+\infty$ , dependendo do intervalo de  $x$ .

#### 2.6.2.4.2 Ajustamento do Modelo de Regressão Logístico Simples

Considerando-se o exposto por HOSMER & LEMESHOW (1989), seja a amostra com  $n$  observações independentes  $(x_i, y_i)$ ,  $i = 1, 2, \dots, n$ , onde  $y_i$  denota o valor de uma variável resposta dicotômica e  $x_i$  é o valor de uma variável independente para o  $i$ -ésimo elemento. A variável resposta é representada pela presença (valor igual a um) ou ausência (valor igual a zero) da resposta. Para ajustar o modelo de regressão logística mostrado na equação (59) para um conjunto de dados, é necessária a estimação dos valores de  $\beta_0$  e  $\beta_1$ , que são parâmetros desconhecidos.

O método comumente empregado para a estimação dos parâmetros é o da máxima verossimilhança. De forma geral, este método obtém os valores para os parâmetros desconhecidos que maximizam a probabilidade de obter o conjunto de dados observados.



A aplicação do método começa com a construção de uma função chamada função de verossimilhança. Essa função expressa a probabilidade conjunta dos dados observados como uma função dos parâmetros desconhecidos. Os estimadores de máxima verossimilhança são escolhidos de forma a maximizar esta função. A seguir, são descritos as etapas necessários a obtenção destes valores.

Se  $Y$  é codificado como zero ou um então a expressão  $\pi(x)$  mostrada na equação (59) fornece (para um valor arbitrário de  $\beta' = (\beta_0, \beta_1)$ , o vetor de parâmetros) a probabilidade condicionada de  $Y$  ser igual a 1 sendo conhecido  $x$ . Esta probabilidade é denotada por  $P(Y = 1 | x)$ . Segue que a quantidade  $1 - \pi(x)$  fornece a probabilidade condicionada que  $Y$  ser igual a zero dado  $x$ , ou seja,  $P(Y = 0 | x)$ . Desse modo esses pares  $(x_i, y_i)$ , se  $y_i = 1$  a contribuição para a função de verossimilhança é  $\pi(x_i)$ , e para aqueles pares onde  $y_i = 0$  a contribuição para a função de verossimilhança é  $1 - \pi(x_i)$ , sendo que  $\pi(x_i)$  é o valor de  $\pi(x)$  calculado para  $x_i$ . De acordo Hosmer e Lemeshow (1989) uma maneira conveniente de expressar a contribuição para a função de verossimilhança referentes aos pares  $(x_i, y_i)$  é:

$$(x_i) = \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (61)$$

Considerando-se as observações independentes, a função de verossimilhança é obtida como o produto dos termos mostrados na expressão 62:

$$\iota(\beta) = \prod_{i=1}^n \zeta(x_i) \quad (62)$$

O princípio para maximizar a verossimilhança é o estimador de  $\beta$  que maximiza a equação (62), e sua solução é obtida com a aplicação do log. O resultado desta aplicação é:

$$L(\beta) = \ln[\iota(\beta)] = \sum_{i=1}^n \{y_i \ln[\pi(x_i)] + (1 - y_i) \ln [1 - \pi(x_i)]\} \quad (63)$$

Para determinar o valor de  $\beta$  que maximiza  $L(\beta)$  em relação a  $\beta_0$  e  $\beta_1$ , definindo os resultados das expressões igual a zero. Estas equações são:

$$\sum_{i=1}^n [y_i - \pi(x_i)] = 0 \quad (64)$$

e

$$\sum_{i=1}^n x_i [y_i - \pi(x_i)] = 0 \quad (65)$$

as expressões (64) e (65) são denominadas equações de verossimilhança (HOSMER; LEMESHOW, 1989).

O valor de  $\beta$  obtido através da solução das equações (64) e (65) é chamado de estimativa de máxima verossimilhança e é denotado por  $\hat{\beta}$ .

Uma consequência da equação (65) é que o somatório dos valores observados de  $y$  é igual ao somatório dos valores preditos ou esperados de  $y$ :

$$\sum_{i=1}^n y_i = \sum_{i=1}^n \hat{\pi}(x_i) \quad (66)$$

#### 2.6.2.4.3 Modelo de Regressão Logístico Múltiplo

Na seção anterior foi apresentada a solução para o modelo logístico considerando-se apenas uma variável independente. Nesta seção é apresentada a formulação para a utilização do modelo logístico múltiplo com  $p$  variáveis independentes (explicativas).

Seja a coleção de  $p$  independentes variáveis que na forma de vetor é:  $\underline{x}' = (x_1, x_2, \dots, x_p)$ . A probabilidade condicionada da resposta é denotada por  $P(Y = 1 | \underline{x}) = \pi(\underline{x})$ . Então a logit do modelo de regressão logístico múltiplo é dado pela equação (HOSMER; LEMESHOW, 1989):

$$g(\underline{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p \quad (67)$$

e

$$\pi(\underline{x}) = \frac{e^{g(\underline{x})}}{1 + e^{g(\underline{x})}} \quad (68)$$

Considerando que se tem  $n$  observações independentes do par  $(x_i, y_i)$ ,  $i = 1, 2, \dots, n$ . Como ocorre no caso univariado, o ajuste do modelo requer que seja obtido o estimador do vetor de parâmetros  $\underline{\beta}' = (\beta_0, \beta_1, \dots, \beta_p)$ . O método de estimação empregado no caso multivariado é o mesmo, ou seja, máxima verossimilhança. A função de verossimilhança é parecida com a equação (62), mudando  $\pi(\underline{x})$ , mostrado na expressão 68. Haverá  $p + 1$  equações de verossimilhança que são obtidas diferenciando a função log verossimilhança em relação a  $p + 1$  coeficientes. As equações de verossimilhança resultantes são expressas por:

$$\sum_{i=1}^n [y_i - \pi(\underline{x}_i)] = 0 \quad (69)$$

e

$$\sum_{i=1}^n x_{ij} [y_i - \pi(\underline{x}_i)] = 0 \quad (70)$$

com  $j = 1, 2, \dots, p$ .

A solução das equações de verossimilhança exige um processo de solução iterativo e  $\underline{\hat{\beta}}$  denota a solução destas equações. Deste modo, os valores ajustados para o modelo de regressão múltiplo são  $\hat{\pi}(x_i)$ , o valor da expressão 67 é obtido empregando-se  $\underline{\hat{\beta}}$  e  $\underline{x}_i$ .

#### 2.6.2.4.4 Modelo de Regressão Logístico Múltiplo com resposta politômica

A regressão logística é mais freqüentemente empregada para modelar a relação entre uma variável resposta dicotômica e um conjunto de covariáveis, mas com umas poucas modificações ela pode ser também usada quando a variável resposta é politômica. Apresenta-se o mesmo desenvolvimento mostrado por

Hosmer e Lemeshow (1989) do modelo logístico múltiplo para uma variável resposta dividida em três categorias.

Considere-se que as categorias da variável resposta,  $Y$ , são codificadas como 0, 1 e 2. Lembrando que o modelo logístico para uma resposta dicotômica parametrizada em termos da logit de  $Y = 1$  versus  $Y = 0$ . Para o caso com três categorias têm-se duas funções logit: uma para  $Y = 1$  versus  $Y = 0$ , e outra para  $Y = 2$  versus  $Y = 0$ . O grupo codificado como  $Y = 0$  será o grupo de referência. A combinação logit  $Y = 2$  versus  $Y = 1$  pode ser obtida através da diferença entre a logit  $Y = 2$  versus  $Y = 0$  e a logit  $Y = 1$  versus  $Y = 0$ .

O vetor  $x$  de  $p$  covariáveis tem tamanho  $p + 1$  com  $x_0 = 1$  considerando-se o termo constante da expressão. As duas funções logit resultam em:

$$g_1(\underline{x}) = \ln \left[ \frac{P(Y = 1|\underline{x})}{P(Y = 0|\underline{x})} \right]$$

$$g_1(x) = \beta_{10} + \beta_{11}x_1 + \beta_{12}x_2 + \dots + \beta_{1p}x_p$$

$$g_1(x) = (1, \underline{x}')\underline{\beta}_1 \quad (71)$$

e ainda

$$g_2(x) = \ln \left[ \frac{P(Y = 2|\underline{x})}{P(Y = 0|\underline{x})} \right]$$

$$g_2(x) = \beta_{20} + \beta_{21}x_1 + \beta_{22}x_2 + \dots + \beta_{2p}x_p$$

$$g_2(x) = (1, \underline{x}')\underline{\beta}_2 \quad (72)$$

Segue que as três probabilidades condicionais referente a cada categoria de resposta dado o vetor de covariáveis são:

$$P(Y = 0|\underline{x}) = \frac{1}{1 + e^{g_1(\underline{x})} + e^{g_2(\underline{x})}} \quad (73)$$

$$(74)$$

$$P(Y = 1|\underline{x}) = \frac{e^{g_1(\underline{x})}}{1 + e^{g_1(\underline{x})} + e^{g_2(\underline{x})}}$$

$$P(Y = 2|\underline{x}) = \frac{e^{g_2(\underline{x})}}{1 + e^{g_1(\underline{x})} + e^{g_2(\underline{x})}} \quad (75)$$

Observando a simbologia empregada para o modelo dicotômico, fazendo  $\pi_j(x) = P(Y = j|x)$  para  $j = 0, 1, 2$  cada um dos quais é função do vetor de  $2(p + 1)$  parâmetros  $\beta' = (\beta_1', \beta_2')$ .

Uma forma geral para a probabilidade condicionada no modelo de três categorias é:

$$P(Y = j|\underline{x}) = \frac{e^{g_j(\underline{x})}}{\sum_{k=0}^2 e^{g_k(\underline{x})}} \quad (76)$$

onde o vetor  $\beta_0 = 0$  desde que  $g_0(x) = 0$ .

A codificação das variáveis ocorre da seguinte maneira:

- se  $Y = 0$  então  $Y_0 = 1, Y_1 = 0$ , e  $Y_2 = 0$ ;
- se  $Y = 1$  então  $Y_0 = 0, Y_1 = 1$ , e  $Y_2 = 0$ ;
- finalmente, se  $Y = 2$  então  $Y_0 = 0, Y_1 = 0$ , e  $Y_2 = 1$ .

Destaca-se que  $\sum Y_j = 1$ .

A função de verossimilhança condicionada para uma amostra de  $n$  observações independentes resulta em:

$$l(\beta) = \prod_{i=1}^n [\pi_0(x_i)^{y_{0i}} \pi_1(x_i)^{y_{1i}} \pi_2(x_i)^{y_{2i}}] \quad (77)$$

Aplicando log a função e valendo-se que  $\sum Y_{ji} = 1$  para cada  $i$ , a função log verossimilhança é

$$L(\beta) = \sum_{i=1}^n y_{1i} g_1(x_i) + y_{2i} g_2(x_i) - \ln(1 + e^{g_1(x_i)} + e^{g_2(x_i)}) \quad (78)$$

As equações de verossimilhança são resolvidas através das derivadas parciais de primeira ordem de  $L(\beta)$  em relação a cada uma das  $2(p + 1)$  variáveis desconhecidas. Uma simplificação da notação empregando  $\pi_{ji} = \pi_j(x_i)$  resulta em:

$$\frac{\delta L(\beta)}{\delta \beta_{jk}} = \sum_{i=1}^n x_{ki}(y_{ji} - \pi_{ji}) \quad (79)$$

para  $j = 1, 2$  e  $k = 0, 1, 2, \dots, p$  e ainda  $x_{0i} = 1$ .

Os estimadores de máxima verossimilhança ( $\hat{\beta}$ ) são obtidos fazendo as equações (79) iguais a zero e resolvendo  $\beta$ . A solução pode ser obtida através do emprego de um método iterativo.

#### 2.6.2.4.2 Modelo de Regressão Logístico Oculto

Se ocorrer a separação total entre os grupos, os Estimadores de Máxima Verossimilhança não existem e com isso não é possível determinar o Modelo Logístico. Neste sentido, o Modelo de Regressão Logístico Oculto Dicotômico apresentado por Rousseeuw e Christmann (2003) pode contornar este problema. Este é um modelo estocástico, pois assume que a verdadeira resposta não é observável e que existe uma variável que pode ser observada que é relacionada com a verdadeira resposta.

Segundo Guimarães (2006), para se aplicar o Modelo de Regressão Logístico Oculto para uma variável com resposta politômica, devem ser consideradas  $k$  variáveis não observáveis  $T_1, T_2, \dots, T_k$ , que tem os valores no intervalo de  $f_i$  ou  $s_i$  com  $i$  variando de 1 até  $k$ . A Figura 10 apresenta este modelo estocástico.

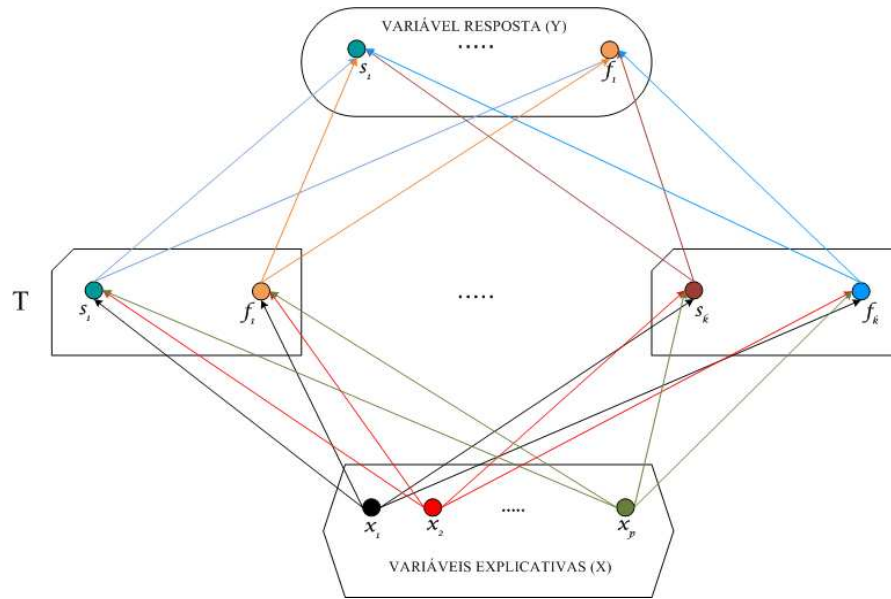


FIGURA 10 - MODELO DE REGRESSÃO LOGÍSTICO OCULTO COM VARIÁVEL RESPOSTA POLITÔMICA  
 FONTE: Adaptado de Guimarães (2006)

O desenvolvimento para a determinação do Modelo de Regressão Logístico Oculto (RLO) com resposta politômica como mostrado na Figura 10 foi desenvolvido por GUIMARÃES (2006) e adiante é apresentado o desenvolvimento utilizado por este autor.

O verdadeiro estado será  $T_j = s_j$  e  $Y = j$  com  $j$  variando de 1 até  $k$ . A probabilidade condicionada é dada por:

$$P(Y = j | T_j = s_j) = \delta s_j \quad (80)$$

A analogia é válida para  $T_j = f_j$  e  $Y = j$

$$P(Y = j | T_j = f_j) = \delta f_j \quad (81)$$

De forma complementar obtêm-se

$$P(Y \neq j | T_j = f_j) = 1 - \delta f_j \quad (82)$$

E para  $\delta f_j$

$$P(Y \neq j | T_j = f_j) = 1 - \delta f_j \quad (83)$$

E considerando então o intervalo para os deltas tem-se:

$$0 < \delta f_j < 0,5 < \delta s_j < 1 \text{ e } j = 1, 2, \dots, k$$

Ainda, segundo Guimarães (2006) existem  $n \times k$  variáveis que não são observáveis ( $T_{ij}$ ), com  $i$  variando de 1 até  $n$  e  $j$  variando de 1 até  $k$ , provenientes de  $k$  vetores ( $\underline{\theta}_1, \underline{\theta}_2, \dots, \underline{\theta}_k$ ) de tal modo que:

$$\underline{\theta}_j = [\underline{\theta}_{j0} \quad \underline{\theta}_{j1} \quad \underline{\theta}_{j2} \quad \dots \quad \underline{\theta}_{jp}]'$$

E com  $\underline{\theta}_k = \underline{0}$ .

O estimador de máxima verossimilhança para  $T_j$  ( $Y = y$ ) é determinado por:

$$\hat{T}_{MLj}(Y \neq j) = f_j$$

$$\hat{T}_{MLj}(Y = j) = s_j \quad (84)$$

Deve-se proceder ao ajuste das pseudo-observações  $\tilde{y}_{ji}$  através do Modelo de Regressão Logística que é:

$$P(G_s | \underline{X}) = \frac{e^{\theta_s}}{\sum_{j=1}^k e^{\theta_j}} \quad (85)$$

Onde:

$$s = 1, 2, 3, \dots, k$$

$$\theta_k = 0$$

$$\theta_s = \theta_{s0} + \theta_{s1}x_1 + \theta_{s2}x_2 + \theta_{s3}x_3 + \dots + \theta_{sp}x_p = \underline{\theta}'_s \underline{X}$$

A função estimada de máxima verossimilhança é determinada por:



$$\iota(\underline{\theta}|\underline{\tilde{Y}}) = \prod_{i=1}^n \prod_{j=1}^k [P(T_j|X_i)]^{\tilde{Y}_{ji}} \quad (86)$$

A função Log-Verossimilhança Estimada é obtida aplicando o logaritmo neperiano, o que resulta em:

$$L(\underline{\theta}|\underline{\tilde{Y}}) = \sum_{i=1}^n [1 + \sum_{j=1}^{k-1} e^{\theta_j}] \quad (87)$$

Os estimadores de  $\theta_j$  são os valores que maximizam a equação (74). Conforme sugerido por GUIMARÃES (2006) empregou-se o Método da Máxima Verossimilhança:

$$\frac{\delta L(\underline{\theta})}{\delta \theta_{jm}} = \sum_{i=1}^n x_{mi} [\tilde{y}_{ji} - P(G_j|\underline{x}_i)] \quad (88)$$

Sendo que:

$$j = 1, 2, \dots, k - 1$$

$$m = 0, 1, 2, \dots, p.$$

Ainda segundo Guimarães (2006), a equação (87) deve ser estritamente côncava. Então, deve-se considerar  $\theta_1$  e  $\theta_2$ , vetores de um subespaço convexo de  $R^{p+1}$ . A função L definida no mesmo subespaço é estritamente convexa, para quaisquer valores de  $\theta_1$  e  $\theta_2$ , com  $\theta_1 \neq \theta_2$  e ainda  $\lambda$ , no intervalo de 0 a 1 se:

$$L(\lambda \underline{\theta}_1 + (1 - \lambda) \underline{\theta}_2) < \lambda L(\underline{\theta}_1) + (1 - \lambda) L(\underline{\theta}_2) \quad (89)$$

Para uma função S, no mesmo subespaço, é estritamente côncava se  $L = -S$  é estritamente convexa.

Sejam  $\theta_1$  e  $\theta_2$ , vetores de um subespaço convexo de  $R^{p+1}$ ,  $\lambda$  variando de 0 até 1 e função:

$$S(\underline{\theta}|\underline{\tilde{Y}}) = -L(\underline{\theta}|\underline{\tilde{Y}})$$

Da equação (73) têm-se:

$$S((\lambda \underline{\theta}_1 + (1 - \lambda) \underline{\theta}_2)|\underline{\tilde{Y}}) \quad (90)$$

Que é igual a:

$$\sum_{i=1}^n \left[ \sum_{j=1}^{k-1} \tilde{y}_{ji} \underline{\theta}'_{2j} \underline{x}_i + \lambda \tilde{y}_{ji} \underline{\theta}'_{1j} \underline{x}_i - \lambda \tilde{y}_{ji} \underline{\theta}'_{2j} \underline{x}_i - \ln \left( 1 + \sum_{j=1}^{k-1} e^{(\lambda \theta'_{1j} + \lambda \theta'_{2j} - \lambda \theta'_{2j})} \right) x_i \right] \quad (91)$$

E ainda:

$$\lambda S(\underline{\theta}_1 | \tilde{Y}) + (1 - \lambda) S(\underline{\theta}_2 | \tilde{Y}) \quad (92)$$

Igual a:

$$- \sum_{i=1}^n \left[ \sum_{j=1}^{k-1} \tilde{y}_{ij} (\theta'_{2j} + \lambda \theta'_{1j} - \lambda \theta'_{2j}) \underline{x}_i - \ln (1 + \sum_{j=1}^{k-1} e^{(\lambda \theta'_{1j} x_i)}) - \ln (1 + \sum_{j=1}^{k-1} e^{(\theta'_{2j} - \lambda \theta'_{2j})} x_i) \right] \quad (93)$$

Guimarães (2006), apresenta a matriz de informação como sendo:

$$I(\underline{\theta}) = \begin{bmatrix} T_{11} & T_{12} & \dots & T_{1(k-1)} \\ T_{21} & T_{22} & \dots & T_{2(k-1)} \\ \vdots & \vdots & \dots & \vdots \\ T_{(k-1)1} & T_{(k-1)2} & \dots & T_{(k-1)(k-1)} \end{bmatrix} \quad (94)$$

Cada elemento da matriz 94 é definido como:

$$T_{ij} = \begin{cases} \underline{X}' [diag[P(G_s | x_i)(1 - P(G_s | x_i))]] \underline{X} & i = j \text{ e } s = i \\ (1 - ) \underline{X}' [diag[P(G_i | x_i)P(G_j | x_i)]] & i \neq j \end{cases} \quad (95)$$

Foram adotados como valores de  $\delta s_j$  e  $\delta f_j$  propostos por Guimarães (2006):

$$\delta s_j = Y_j$$

e

$$\delta f_j = 1 - Y_j$$

Para  $Y_j = 0,0001$ .

### 3 MATERIAL E MÉTODOS

#### 3.1 LOCALIZAÇÃO DA ÁREA DE ESTUDO

O local escolhido para o desenvolvimento dos trabalhos situa-se no município de União da Vitória, sul do estado do Paraná. É uma transição do ambiente urbano para o rural, caracterizada pela alternância na cobertura do solo entre áreas agrícolas, florestais (remanescentes e plantadas) e mancha urbana. A Figura 11 apresenta o mapa do Estado do Paraná e a localização da área de estudo no município.

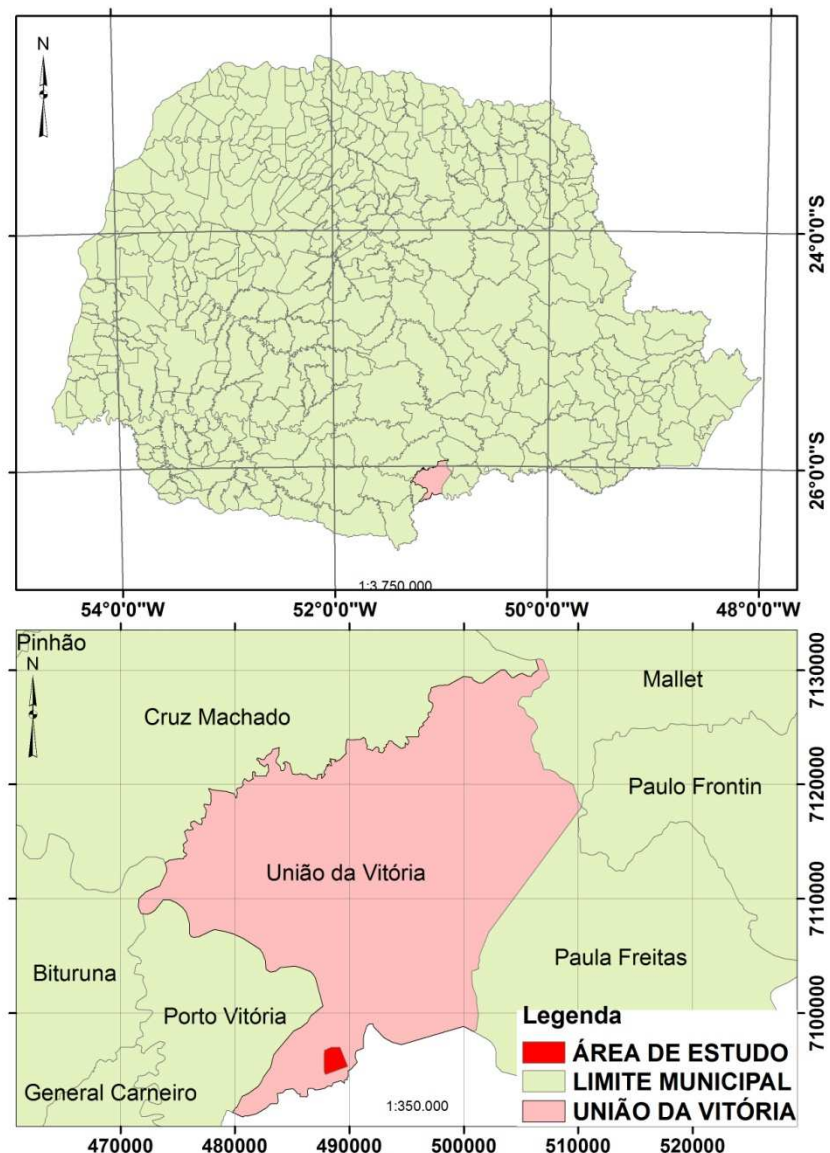


FIGURA 11 – LOCALIZAÇÃO DA ÁREA DE ESTUDO

Como critérios para seleção deste local citam-se a presença das diferentes formações vegetais, a existência de um perfilamento a LASER e de um recobrimento aéreo.

### 3.2 MATERIAIS

Para a realização desta pesquisa empregaram-se os seguintes materiais, programas e informações:

- 01 (um) computador pessoal, com processador Pentium Core 2 Duo 2.66 Ghz, 1.333 Mhz de barramento, 8 Gb de memória RAM, placa de vídeo GFORCE 8600, com 512 Mb de memória, disco rígido de 500 Gb de capacidade de armazenagem, monitor LCD de 19" e kit multimídia;
- 01 (uma) impressora jato de tinta, marca HP modelo Deskjet 3920;
- 01 (um) par de receptores de sinais GPS marca Ashtech, modelo Promark 2 com acessórios;
- 01 (um) veículo para os deslocamentos de campo;
- 01 (uma) câmara fotográfica digital;
- 02 (dois) arquivos digitais (primeiro e último retorno) contendo os dados brutos gerados pelo ALS OPTech / INS GPS APPLANIX, pertencente ao Instituto de Tecnologia para o Desenvolvimento (LACTEC - UFPR). O sistema de referência das coordenadas para o rastreamento foi o WGS84 (*World Geodetic System*, 1984);
- 01(uma) imagem de intensidade gerada a partir dos dados brutos;
- 03 (três) fotografias aéreas verticais preto e branco, em escala 1/10.000, em formato digital;
- 01 (um) estereoscópio de espelhos, marca Leica e oculares;
- 01(um) restituidor B8-S, marca Wild;
- 01 (uma) cópia do programa Microstation V8 8.1;
- 01 (uma) cópia do programa Matlab 2007;
- 01 (uma) cópia do programa Surfer 8.0;
- 01 (uma) cópia do programa Ashtech Solutions 2.6;
- 01 (uma) cópia do programa Arcgis 9.2;
- 01 (uma) cópia do programa Terrascan;

- Folhas A4, transparências, caneta para retroprojeter 1.0 mm, cartuchos de tinta, prancheta, fita adesiva, grafite, lapiseira e borracha.

As fotografias aéreas foram convertidas do formato analógico para o digital empregando-se um *scanner* fotogramétrico, com resolução geométrica de 1.024 dpi.

O levantamento por varredura LASER empregou o sistema OPTECH ALTM (*Airborne LASER Terrain Mapper*) 2050 /INS GPS APPLANIX, pertencente ao Instituto de Tecnologia para o Desenvolvimento (LACTEC - UFPR). Este recobrimento foi realizado em 13 de novembro de 2003 a uma altura aproximada de 1.000 metros. As varreduras foram realizadas no formato zig-zag, transversal à linha de voo.

Foram disponibilizados dois arquivos digitais, relativos ao primeiro e último retorno, no formato de texto com um ponto por linha. Os arquivos contêm informações sobre: coordenadas espaciais Leste (E), Norte (N), altitude ortométrica (h) e intensidade de refletância (I).

O recobrimento aéreo fotogramétrico ocorreu em 2 de junho de 2004, oito meses após a varredura LASER. Foi empregada uma câmara métrica RC 10 Wild, pertencente ao Departamento de Geomática da Universidade Federal do Paraná. A distância focal calibrada da câmara é de 153,172 mm, sendo que as fotografias foram tomadas a uma altura média do solo de 1.530 metros o que resultou em uma escala média de 1/10.000. Foram obtidas neste recobrimento 86 aerofotos que recobriram uma área de 183,53 km<sup>2</sup>.

O limite da varredura LASER foi sobreposto ao foto-índice e então se escolheram as três fotografias aéreas abrangidas por este limite e de interesse para a pesquisa.

A disposição do limite da varredura LIDAR e do foto-índice para a área do experimentos são apresentados na Figura 12.

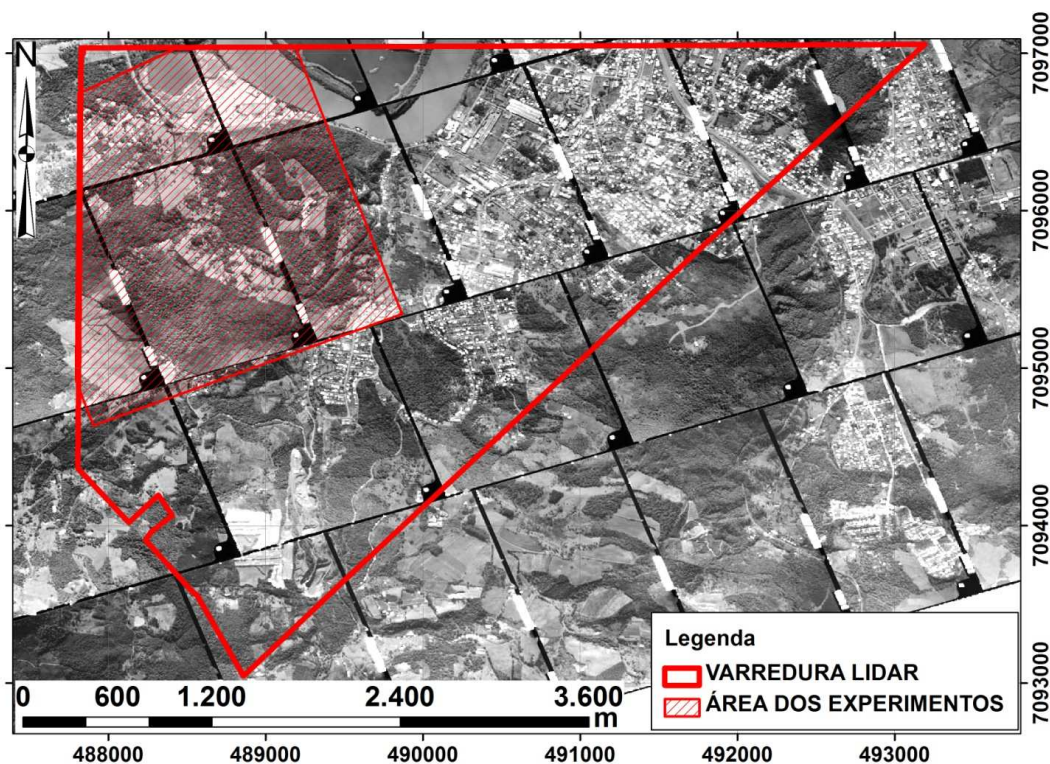


FIGURA 12 – LIMITE DA VARREDURA LASER E DA ÁREA DOS EXPERIMENTOS

A área abrangida pela varredura LASER é de 11,50 Km<sup>2</sup> e contém 22.452.524 pontos, correspondentes ao primeiro e último retorno. Procedeu-se a seleção dos pontos de varredura contidos na área dos experimentos (vide Figura 12) que têm 3,58 Km<sup>2</sup>. O Quadro 1 apresenta a quantidade de pontos selecionados.

	Total de Pontos LASER	Pontos Selecionados	Percentual (%)
Primeiro Retorno	7.836.065	2.265.384	34,90
Último Retorno	14.616.459	4.225.576	65,10
Totais	22.452.524	6.490.960	100,00

QUADRO 1 - QUANTIDADE DE PONTOS LIDAR SELECIONADOS

Do total selecionado o percentual de pontos LIDAR relativos ao primeiro retorno corresponde a 34,90% e 65,10% é o percentual dos pontos referentes ao último retorno.

A densidade média de pontos para a área dos experimentos é de 1,81 ponto/m<sup>2</sup> ou ainda, expressando linearmente este valor, 1,35 ponto/m.

### 3.3 METODOLOGIA DE EXECUÇÃO DA PESQUISA

As atividades da pesquisa necessárias ao mapeamento da cobertura do solo a partir da grade LASER foram divididas nas seguintes etapas:

- Fotointerpretação das fotografias aéreas verticais;
- Reambulação e obtenção do apoio fotogramétrico;
- Referenciamento cartográfico da fotointerpretação;
- Criação das grades regulares para armazenamento dos pontos LASER;
- Determinação do Modelo Digital do Terreno;
- Determinação da altura da cobertura vegetal;
- Obtenção das amostras referentes a cobertura do solo;
- Determinação da cobertura vegetal através da classificação das grades regulares e
- Análise do acerto dos classificadores empregados.

As seções a seguir detalham a metodologia empregada para a obtenção dos resultados.

#### 3.3.1 Fotointerpretação das Fotografias Aéreas Verticais

Para verificar a exatidão da metodologia proposta nesta pesquisa, os dados gerados a partir das observações fotogramétricas, associados com os obtidos na reambulação foram considerados, neste trabalho, como verdade terrestre terrestre.

Desta forma, para que a espacialização das observações fosse o mais próxima possível da realidade de campo foi necessário o treinamento na fotointerpretação da cobertura do solo. Este treinamento na fotointerpretação da vegetação foi feito com auxílio dos professores do Departamento de Ciências Florestais da Universidade Federal do Paraná.

Selecionaram-se dois pares estereoscópicos formados pelas fotografias 02, 03 e 04 da faixa 03, resultando nos modelos mo0203 e mo0304.

As Figuras 13 e 14 apresentam os dois modelos estereoscópicos empregados na pesquisa. O método de visualização apresentado é o anaglifo, e a



visualização em três dimensões é possível com o emprego dos óculos de anaglifo que estão à disposição na última página deste trabalho.



FIGURA 13 – MODELO ESTEREOSCÓPICO 0203

A análise da Figura 13 destaca a predominância de um ambiente rural, com uma variedade de formações vegetais incluindo áreas de florestas naturais em diferentes estágios, florestas plantadas, campos e agropecuária. Ao norte percebe-se a existência de um trecho da mancha urbana de União da Vitória, bem como, cavas de areia.





FIGURA 14 – MODELO ESTEREOSCÓPICO 0304

Na Figura 14, nota-se a predominância do ambiente rural, contudo a mancha urbana está mais visível na direita do modelo (área alagável). Ao norte é possível visualizar o rio Iguaçu e cavas de areia.

O relevo do local apresenta altitudes variando de 740 a 965 metros e as declividades variando de 0 a 34°. Os locais com valores elevados de declividade estão ocupados por formações naturais em estágio de sucessão mais avançados.

Para a interpretação dos modelos estereoscópicos empregou-se a Fotointerpretação Direta e a Correlativa. A Fotointerpretação Direta trata dos objetos visíveis nas aerofotos. É a identificação de elementos culturais e naturais tais como: edificações, obras de engenharia, campos cultivados, rios, florestas e plantios agrícolas. A Fotointerpretação além da análise direta inclui um exame detalhado e uma cuidadosa avaliação de todos os outros elementos na fotografia aérea sendo possível identificar elementos não visíveis, como os estágios de sucessão das formações naturais (DISPERATI *et al.*, 2007).

Neste contexto, todos os objetos visíveis nas aerofotos possuem características próprias que auxiliam o processo de identificação. As características inerentes aos objetos e que foram empregadas na fotointerpretação são: tonalidade, tamanho, forma, sombra, textura, padrão, local, associação e tempo. A seguir são descritas as características dos objetos empregadas como parâmetros de classificação, ainda de acordo com a visão de Disperati *et al.* (2007).

A tonalidade é fundamental, sendo expressa em tons de cinza, variando do preto ao branco. É o único elemento de reconhecimento observado diretamente e o que mais fornece informações ao intérprete. Os tons de cores dependerão de várias características, mas especialmente, das características espectrais dos objetos, das condições no instante da tomada da fotografia, do material fotográfico, do processamento do filme e seus derivados.

O tamanho com que os objetos aparecem é função da escala média da fotografia, sendo uma referência valiosa para a sua identificação.

A forma diz respeito à ocorrência bidimensional do objeto estudado. As obras feitas pelo homem (culturais) têm formas uniformes com traçados em linhas retas, o que é pouco comum nos objetos naturais.

A sombra é consequência da forma do objeto e da hora da tomada da fotografia. Este parâmetro pode prejudicar a fotointerpretação, uma vez que, a região de sombra não apresenta nitidez dos objetos. As fotografias empregadas foram obtidas próximo das 13 horas, o que diminuiu bastante este parâmetro.

A textura é estabelecida pela repetição da tonalidade dentro do objeto ou dentro de grupos de objetos que são muito pequenos para serem caracterizados

individualmente. A textura varia de lisa a rugosa dependendo das características dos objetos, da resolução espacial da imagem e da escala da aerofoto.

A repetição de formas ou arranjos de objetos ou suas relações constituem padrões naturais ou culturais. Exemplos destes padrões são as manchas urbanas e a drenagem.

O reconhecimento local implica na relação dos objetos e fenômenos com as respectivas posições geográficas ou as condições do terreno. O conhecimento da localização geográfica é um referencial importante para a identificação dos objetos.

No caso, as cavas existentes no local poderiam ser confundidas com lagos artificiais, o que foi contornado pela inspeção de campo e também conhecimento prévio da existência de uma empresa de extração de areia.

Com relação às formações naturais, a tipologia florestal do local está inserida no bioma Mata Atlântica, sendo classificada como Floresta Ombrófila Mista. Na abrangência da bacia hidrográfica do rio Canoas, nas escarpas da Serra Geral e no Planalto Meridional Brasileiro tem uma situação de ocupação mista, caracterizada pela Região Fito ecológica da Floresta Ombrófila Mista e pela Região Fito ecológica da Estepe (denominada como Campos do Planalto).

A Floresta Ombrófila Mista (FOM) caracteriza-se pela ocorrência predominante do pinheiro-brasileiro ou pinheiro do paraná (*Araucaria angustifolia*) também conhecida como mata de araucária, pinheiral ou pinhal. É um tipo de vegetação típico e exclusivo dos planaltos da região Sul do Brasil, com disjunções na região Sudeste e em países vizinhos (Argentina e Paraguai) (RODERJAN *et al.*, 1993).

Outro parâmetro empregado na fotointerpretação foi a associação, que se refere às adjacências ou convergências de evidências ou correlação de aspectos associados. Por exemplo, a partir da análise de adjacência é possível identificar se uma floresta sofreu alterações humanas.

O tempo refere-se à relação temporal dos fenômenos e objetos. Por exemplo, uma floresta decidual no inverno não apresentará cobertura de folhas, situação oposta à verificada no verão.

Em função dos parâmetros apresentados e da variedade de classes temáticas no local de estudo, optou-se pela definição das chaves de fotointerpretação.

O sucesso da fotointerpretação depende do uso de um guia, o qual ajuda o foto intérprete a identificar rapidamente as características fotográficas. Essas chaves são baseadas em descrições e ilustrações típicas de objetos de uma determinada categoria, por exemplo, indústrias. São organizadas para um estudo comparativo, o intérprete seleciona os diferentes exemplos que mais coincidem com a característica que precisa ser identificada. O procedimento é lento e deve ser feito cuidadosamente, seguindo passo a passo a seqüência correta de trabalho, partindo do geral para o específico.

Com auxílio de um estereoscópio de espelhos foi obtida a visão tridimensional, para o modelo 0203 e para o modelo 0304. A Figura 15 apresenta o modelo 0203 ajustado sob o estereoscópio.

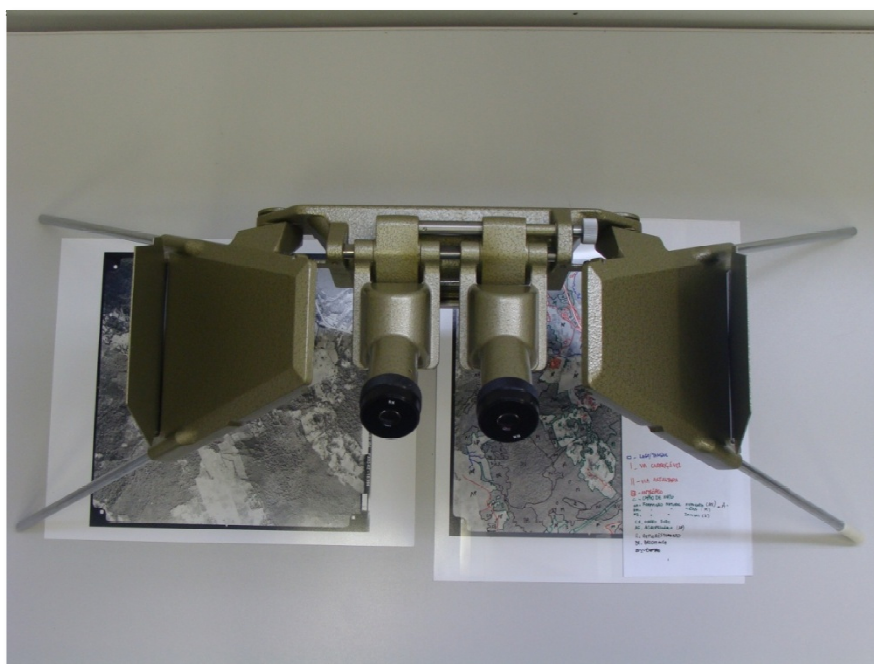


FIGURA 15 – PAR FOTOGRÁFICO AJUSTADO SOB O ESTEREOSCÓPIO DE ESPELHOS

Tomando por referência a classificação vegetal proposta por Veloso, Filho e Lima (1991) e com o auxílio do estereoscópio de espelhos identificaram-se doze classes temáticas referentes à cobertura do solo na área dos experimentos. Estas classes temáticas podem ser agrupadas em:

- Região da Floresta Ombrófila Mista:

Estágio Inicial da Sucessão Vegetal herbáceo-arbustiva - capoeirinha;  
Estágio Inicial da Sucessão Vegetal arbórea - bracatinga;  
Estágio Inicial da Sucessão Vegetal arbórea - capoeira;  
Estágio Médio da Sucessão Vegetal arbórea - capoeirão;  
Estágio Avançado da Sucessão Vegetal - floresta e  
Floresta de galeria.

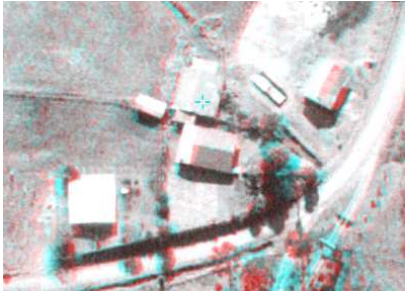

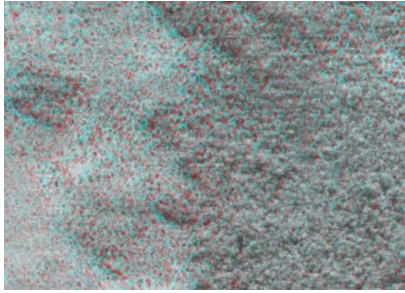
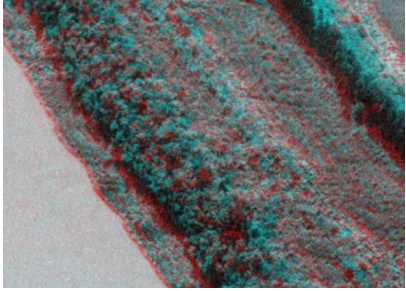
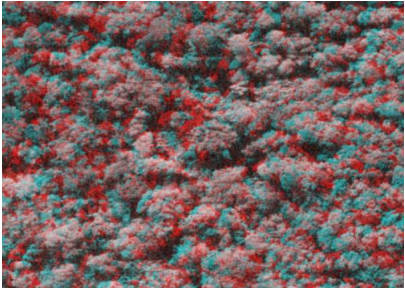
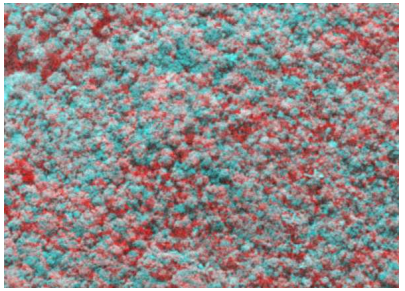
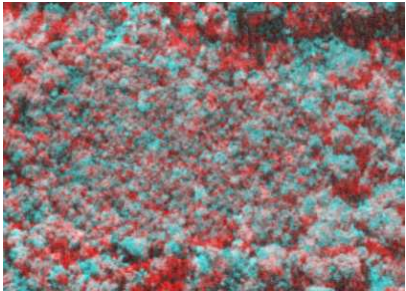
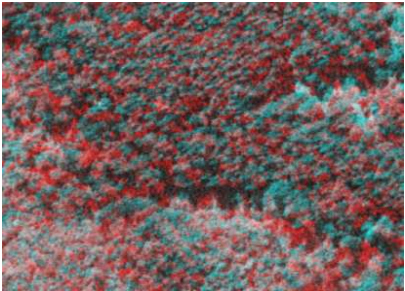

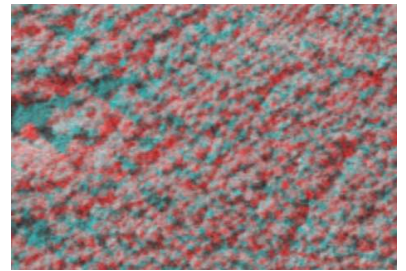
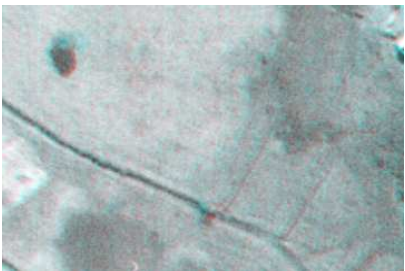

- Áreas de Formações Pioneiras com influência Flúvio-lacustre - várzea.

- Áreas de antropismos:

Agropecuária;  
Reflorestamento;  
Edificações;  
Solo exposto e  
Corpos d'água.

Apresenta-se no quadro 2 a chave da fotointerpretação identificada para cada classe temática na área dos experimentos. A chave está apresentada em anaglifo para permitir a sua visualização tridimensional.



Edificações	Agropecuária	Capoeirinha
		
Floresta de Galeria	Floresta	Capoeirão
		
Capoeira	Reflorestamento	Solo Exposto
		
Bracatinga	Várzea	Corpos D'água
		

QUADRO 2 - CHAVES DE FOTOINTERPRETAÇÃO SELECIONADAS

Como pode ser observado no Quadro 2, além das formações vegetais optou-se por selecionar as classes temáticas referentes à agropecuária, solo exposto, edificações e Corpos d'água. Desta forma, tornou-se possível analisar o acerto das funções discriminantes na classificação destas categorias e verificar a viabilidade do seu emprego para a separação destas classes temáticas.

Com a definição das chaves de fotointerpretação e com o treinamento interpretativo concluído, procedeu-se o delineamento manual das classes em cada um dos modelos. Nesta etapa empregaram-se canetas de retroprojeto coloridas e folhas de material transparente, além das fotografias analógicas e o estereoscópio de espelhos com oculares.

Em cada modelo foi fixada a folha transparente na fotografia da direita e com auxílio do estereoscópio foram delineadas as classes temáticas previamente identificadas. A Figura 16 apresenta o modelo 0203 e a delimitação manual das classes temáticas obtida com a fotointerpretação manual.

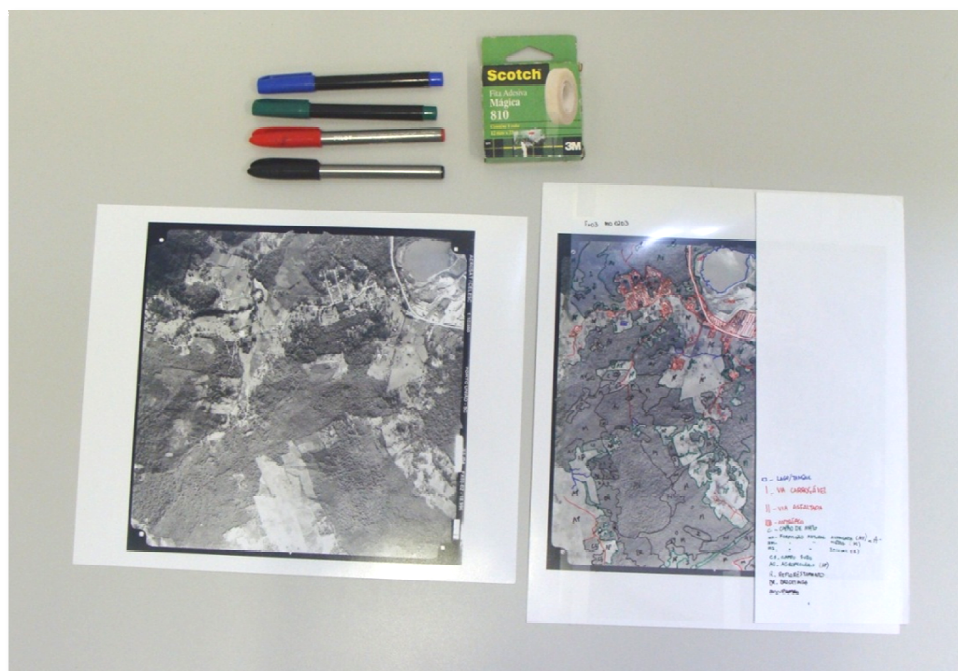


FIGURA 16 – MODELO 0203 COM AS CLASSES TEMÁTICAS DELIMITADAS

Concluídos os trabalhos de fotointerpretação foram realizadas as atividades de reambulação de campo e paralelamente a obtenção do apoio fotogramétrico.

### 3.3.2 Reambulação e Obtenção do Apoio Fotogramétrico

Para a obtenção da base cartográfica digital contendo os elementos foto interpretados realizou-se a restituição fotogramétrica estereoscópica, com a utilização de um restituidor analógico com captura dos dados da representação vetorial em ambiente digital. O estágio inicial nesta atividade consiste na obtenção de modelos estereoscópicos fotogramétricos orientados em relação ao espaço objeto. Para o estabelecimento desta relação é necessária a obtenção das coordenadas tridimensionais de pontos foto-identificáveis (pontos de apoio), referidos a um referencial terrestre.

A finalidade primordial, mas não a única, dos pontos de apoio é a materialização do referencial com que se deseja trabalhar. A fim de materializar um referencial tridimensional sete parâmetros terão de ser determinados, sendo que estes parâmetros definem a posição da origem do sistema, a orientação de seus três eixos e a escala (ANDRADE, 2003).

Segundo Mitishita (1997), a orientação absoluta do modelo fotogramétrico consiste na obtenção do modelo matemático que transformará as coordenadas do referencial de modelo para o referencial geodésico local.

Para tornar possível a transformação mencionada fez-se necessário um conjunto de pontos fotogramétricos com coordenadas conhecidas em ambos os referenciais (pontos foto identificáveis).

Ainda de acordo com Mitishita (1997) considerando que as diferenças físicas entre os referenciais cartesianos ortogonais sejam devidas a rotações e translações nos três eixos coordenados e a um fator de escala, o modelo isogonal tem condições de parametrizar tais diferenças.

O modelo isogonal é expresso por:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \lambda M \begin{bmatrix} X - X_0 \\ Y - Y_0 \\ Z - Z_0 \end{bmatrix} \quad (96)$$

onde:

$[X \ Y \ Z]'$  - Coordenadas de pontos fotogramétricos no referencial geodésico local,



$[x \ y \ z]'$  - Coordenadas de pontos fotogramétricos no referencial de modelo,

$\lambda$  - fator de escala,

$[X_0 \ Y_0 \ Z_0]'$  - Parâmetros de translação e

M - Matriz de rotação ortogonal, proveniente das rotações nos eixos cartesianos.

A Figura 17 apresenta a distribuição dos pontos de apoio empregada na orientação dos modelos. Esta distribuição foi escolhida pois é aproximadamente coincidente com os pontos de Von Gruber que possibilitam a orientação relativa e absoluta do modelo fotogramétrico.

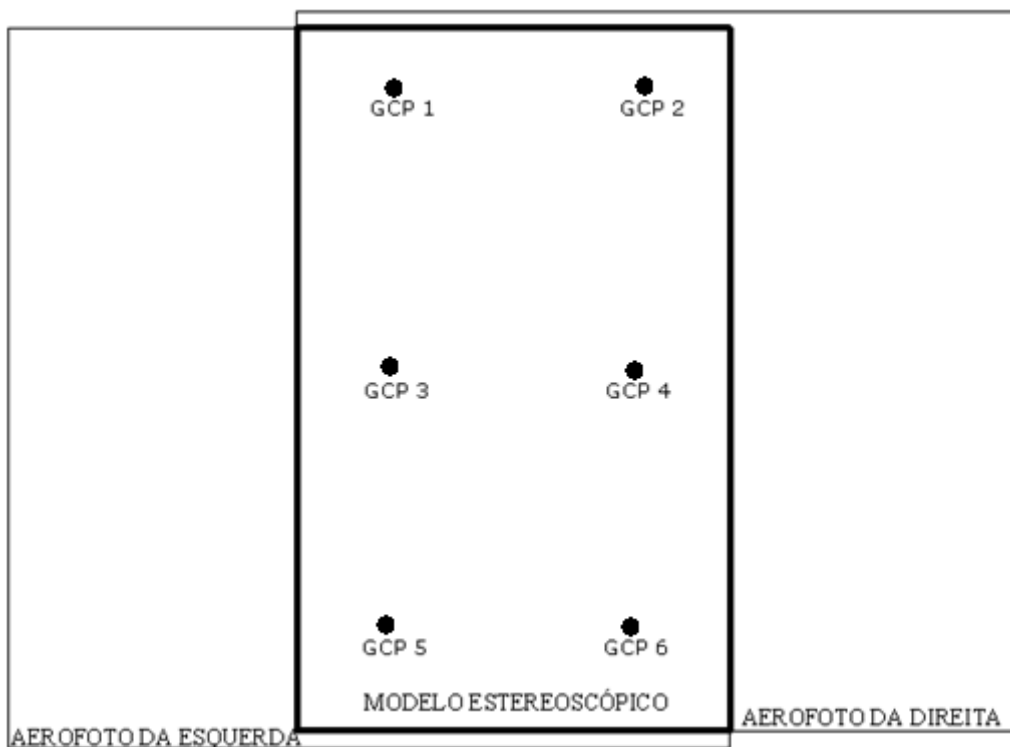


FIGURA 17 – DISTRIBUIÇÃO DOS PONTOS DE APOIO (GCP - *Ground control point*) NO MODELO FOTOGRAMÉTRICO

Com a configuração de pontos mostrada na Figura 17 obtêm-se um sistema de 18 equações e com 7 parâmetros, caracterizando um sistema com super abundância de observações e passível de ajustamento. De acordo com Mitishita (1997), a técnica de ajustamento paramétrico com injunções posicionais controladas com peso é a mais apropriada. Após a determinação dos parâmetros, as coordenadas fotogramétricas obtidas no modelo poderão ser transformadas para o referencial geodésico local com a transformação inversa dada por:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \lambda^{-1} M^T \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \quad (97)$$

A Figura 18 mostra a disposição final dos 15 pontos de apoio nas aerofotos utilizadas neste trabalho.

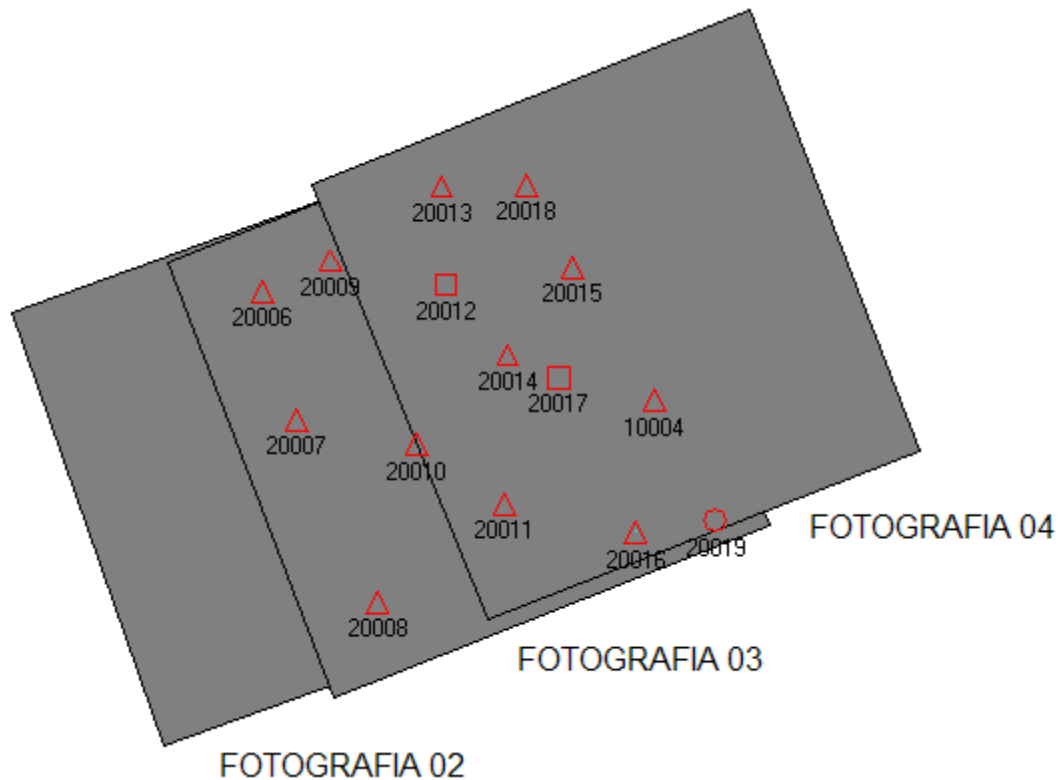


FIGURA 18 – DISTRIBUIÇÃO DOS PONTOS DE APOIO NAS FOTOGRAFIAS

Como pontos de apoio foram selecionados pontos foto-identificáveis, tais como cantos de cerca/muro, intersecções de vias e bordas de edificações.

Para a determinação do referencial dos pontos no terreno foi empregada a Geodésia Celeste através do rastreo das observáveis GPS. Para o rastreo foram adotados os seguintes parâmetros:

- Método de rastreo: diferencial estático rápido;
- Ângulo de corte: 15°;
- Taxa de gravação: 1 segundo;
- Tempo de ocupação por ponto: 20 minutos.
- Receptores: capacidade de rastreo e gravação da onda portadora L1 e do código C/A.

Para alcançar as precisões necessárias à determinação do referencial geodésico optou-se por desenvolver o levantamento com linha de base curta. Empregou-se como ponto base, aquele materializado nas dependências da Universidade do Contestado em Porto União. Desta forma a linha de base não ultrapassou os 6 quilômetros.

As coordenadas UTM (Universal Transversa de Mercator, Meridiano Central 51°w) do ponto base referidas ao Datum WGS84 são:

E - 491.664,821 m / - 0,025 m

N - 7.096.969,217 m / - 0,018 m e

Z - 765,852 m / - 0,072 m.

Sendo que os valores , e referem-se às precisões do ponto base obtidas com transporte das coordenadas em relação ao Sistema Geodésico Brasileiro (SGB).

Com relação ao rastreo dos pontos foto-identificáveis a Figura 19 apresenta um exemplo da execução desta tarefa.

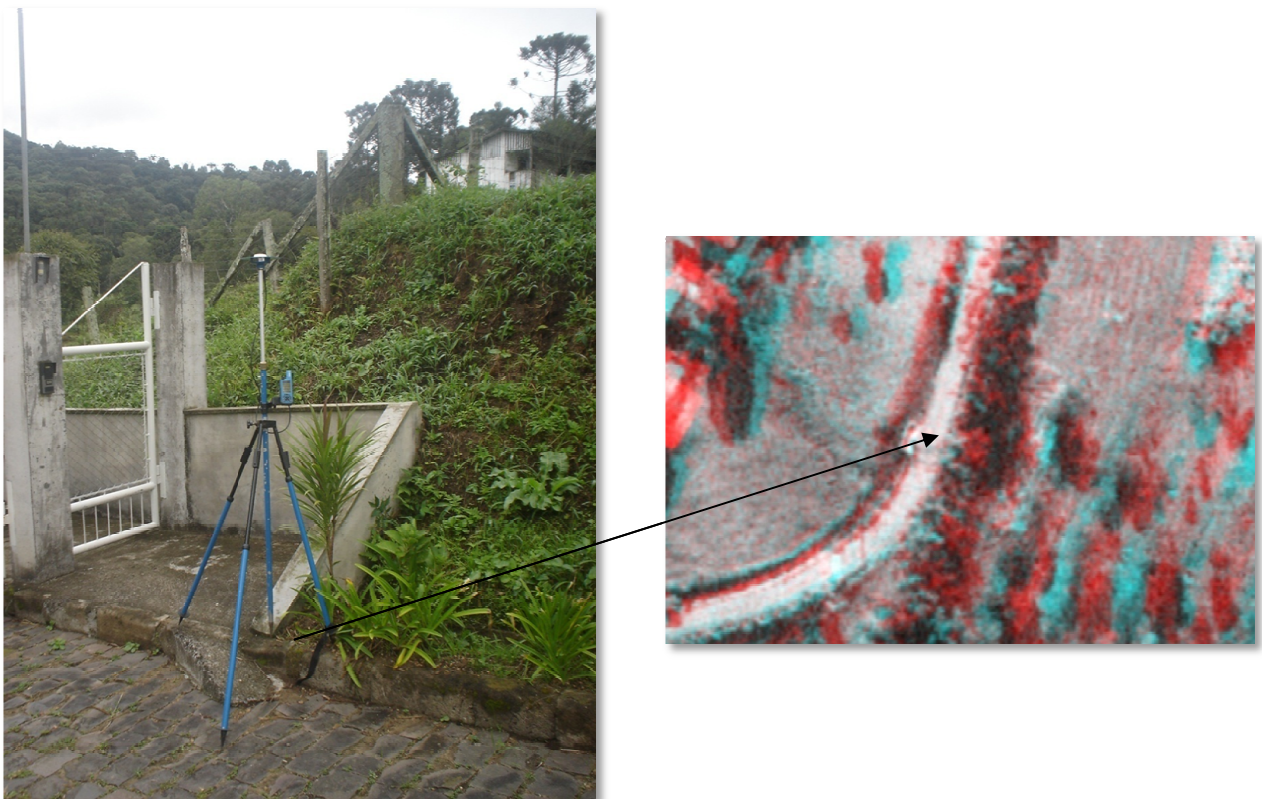


FIGURA 19 – RASTREIO DE UM PONTO FOTO-IDENTIFICÁVEL











Ao todo, foram rastreados 15 pontos foto-identificáveis cuja distribuição pode ser verificada na Figura 18. A coleta das observáveis ocorreu nos dias 25 e 26 de março de 2008 e após a sua conclusão, os dados coletados foram processados no programa Ashtech Solutions 2.6.

Paralelamente à atividade de coleta das observáveis GPS, procedeu-se a reambulação da fotointerpretação. A finalidade da reambulação foi validar e quando necessário, corrigir o mapeamento do uso do solo obtido através da fotointerpretação manual. Esta atividade foi implementada através de inspeções *in locu*, auxiliadas por um engenheiro florestal.

As fotografias aéreas foram empregadas para a orientação no campo, sendo que as inspeções ocorreram para os locais nos quais fosse possível o acesso através de estradas. Para a análise das formações naturais distante das estradas, empregou-se binóculo para auxiliar a visualização e identificação dos fragmentos florestais.

Nos locais próximos às estradas e ocupados por formações vegetais a inspeção ocorria no interior da formação, sendo que o engenheiro florestal determinava as espécies predominantes e com auxílio de um hipsômetro determinava a altura da cobertura vegetal. Desta maneira definia-se o estágio de sucessão vegetal natural: inicial, média e avançada.

A área de estudo foi percorrida e desta forma foi possível validar os resultados da fotointerpretação. O Quadro 3 apresenta imagens das classes inspecionadas em campo.

<p>Agropecuária</p> 	<p>Capoeirinha</p> 
<p>Várzea</p> 	<p>Corpos D'água</p> 
<p>Floresta de Galeria</p> 	<p>Floresta</p> 
<p>Capoeirão</p> 	<p>Capoeira</p> 
<p>Reflorestamento com <i>Pinnus</i> sp.</p> 	<p>Edificações</p> 

QUADRO 3 - CLASSES INSPECIONADAS EM CAMPO

Neste ponto, cabe ressaltar que Conselho Nacional do Meio Ambiente (CONAMA) preconiza a realização do inventário florestal para a determinação da sucessão da vegetação natural a fim de determinar se o estágio é avançado, médio ou inicial. As atividades realizadas em campo, apesar de contarem com o apoio de um engenheiro florestal, caracterizaram-se pela inspeção visual das formações vegetais e, portanto, não satisfazem a orientação do CONAMA.

A área ocupada pela formação vegetal bracatinga foi suprimida e na época da realização das inspeções de campo este local era ocupado por reflorestamento com *Pinnus sp.*. Destaca-se ainda que não foram apresentadas no quadro 2 as fotografias referentes as áreas de solo exposto e vias que foram devidamente identificadas pelas atividades de fotointerpretação e de restituição.

### 3.3.3 Referenciamento Cartográfico da Fotointerpretação

Conforme mencionado, para a representação cartográfica da fotointerpretação, em meio digital, foram empregadas técnicas de restituição fotogramétrica. A restituição planialtimétrica dos dois modelos (0203 e 0304) foi conduzida por profissional experiente em atividades desta natureza. A coleta das informações foi realizada em um restituidor marca Wild, modelo B8S, adaptado para a coleta das informações diretamente no computador (restituidor semi-analítico).

As etapas de trabalho de restituição foram:

- orientação interior: obtida através da coincidência das marcas fiducias da câmera com as correspondentes marcas no restituidor. É necessário também o ajuste da distância principal do restituidor com a da câmera fotogramétrica;
- orientação relativa: nesta etapa procede-se a orientação da posição e atitude do par de fotografias entre si, isto é, sem levar em consideração o referencial terrestre. "Se cinco pares de raios homólogos cruzam-se, todos os demais se cruzarão", teorema de Von Gruber citado por Andrade (2003);
- orientação absoluta: esta etapa é o conjunto de operações necessárias para colocar o modelo estereoscópico, obtido pela orientação relativa referenciado a um sistema de coordenadas geodésicas. Esta operação é feita através do nivelamento, escala e orientação do modelo. Na sequência, o modelo foi

matematicamente orientado em relação ao referencial de coordenadas geodésicas. Para tanto, os parâmetros de translação, rotação e escala são calculados em função das leituras no modelo, de coordenadas de máquina, dos pontos de apoio. Então, a transformação resultante pode ser aplicada a todas as coordenadas de máquina (ANDRADE, 2003);

- Restituição planimétrica: nesta etapa o fotogrametrista utilizando o restituidor semi-analítico procedeu a coleta das observações, nos modelos fotogramétricos, que possibilitou a delimitação da cobertura vegetal. Com auxílio da marca flutuante, foi procedido o registro em meio digital das coordenadas dos polígonos definidores da cobertura vegetal, sendo esta atividade orientada pelas informações da fotointerpretação manual e reambulação. O resultado desta etapa foi o mapa temático referente às classes de cobertura vegetal que foi empregado para avaliar os resultados obtidos pelas funções discriminantes;

- Perfilamento altimétrico: esta atividade almejou a determinação da superfície topográfica (MDT) que foi empregada para avaliar a eficiência do filtro morfológico implementado na pesquisa. Inicialmente, o fotogrametrista definiu perfis altimétricos que recobrem o modelo e em cada perfil foram medidas, com auxílio da marca flutuante do equipamento B8S, as coordenadas tridimensionais de cada ponto cotado. A distância entre perfis e pontos cotados foi definida como sendo 10 metros. Destaca-se ainda que, para os locais cobertos por formações vegetais onde não foi possível visualizar o solo, a altitude dos pontos foi mensurada em função da altura da vegetação verificada nas bordas destas formações.

Para a realização da restituição planialtimétrica da área de estudo foram necessárias 45 horas de trabalho.

### 3.3.4 Criação das Grades Regulares para o Armazenamento dos Pontos LASER

Para otimizar o processamento dos dados LASER optou-se por armazenar todos os pontos em uma grade regular, conforme metodologia adotada por Machado (2006). Esta grade é uma hipermatriz espacialmente definida, que neste estudo foi delimitada pelas seguintes coordenadas (em metros):

- Canto superior esquerdo: E = 487.790, N = 7.097.100;



- Canto inferior direito: E = 489.890, N = 7.094.600.

A Figura 20 mostra retângulo definidor da grade regular sobreposto à área dos experimentos considerada nesta tese.

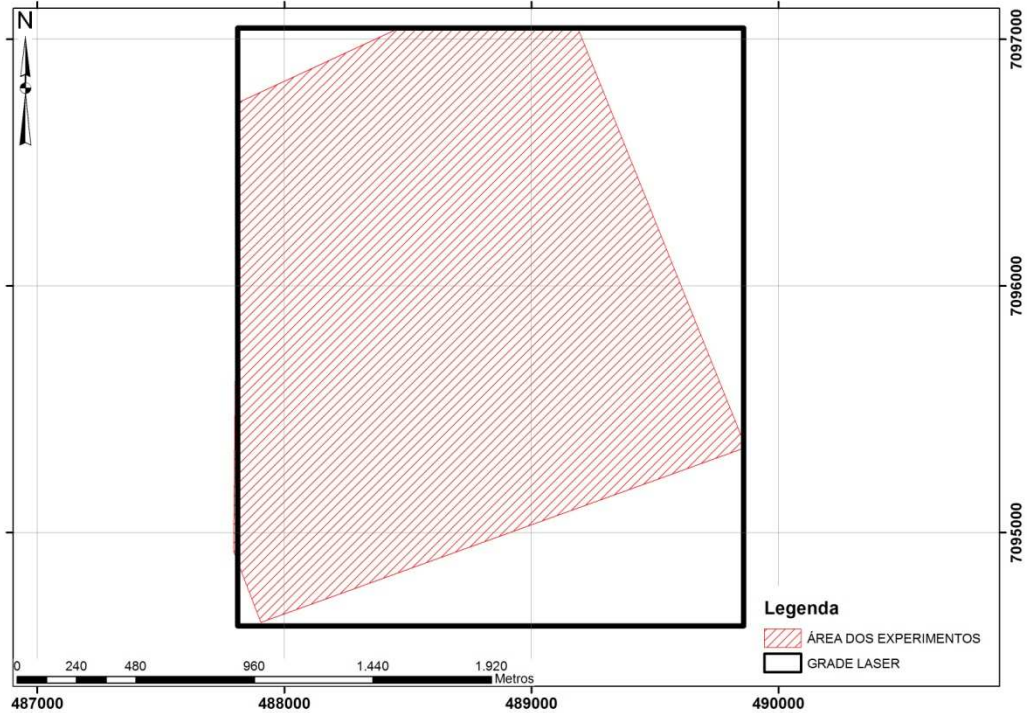


FIGURA 20 – DISPOSIÇÃO DA GRADE LASER E A ÁREA DOS EXPERIMENTOS

No interior de cada célula da grade regular foram alocados os pontos LASER. O critério para alocação dos pontos nas células foi baseado em sua posição bidimensional (E, N):

$$\Delta E = E_i - E_{min} \quad (98)$$

$$\Delta N = N_{max} - N_i \quad (99)$$

Onde:

$\Delta E$  - variação da coordenada E;

$\Delta N$  - variação da coordenada N;

$E_i, N_i$  - coordenadas do ponto analisado;

$E_{min} = 487.790$  metros;



$N_{max} = 7.097.100$  metros.

A célula (linha, coluna) na qual o ponto foi alocado é determinada pela parte inteira do resultado obtido por:

$$\text{linha} = \frac{\Delta N}{r} \quad (100)$$

$$\text{coluna} = \frac{\Delta E}{r} \quad (101)$$

Sendo  $r$  a resolução espacial da grade regular.

Foram escolhidos três valores de  $r$ , resultando três grades regulares. Os valores escolhidos foram:

a) 10 metros: esta resolução foi escolhida de forma a coincidir com o espaçamento dos perfis altimétricos derivados pela fotogrametria e assim facilitar a comparação com os resultados obtidos pelo filtro morfológico adaptado. Ademais os autores Zaksek e Pfeifer (2006) sugerem este espaçamento de grade para área ocupadas por formações naturais mais densas;

b) 5 metros e 1 metro: estas resoluções foram escolhidas visando analisar o efeito de bordas na classificação da cobertura vegetal a partir da grade regular. Ainda cita-se como critério da seleção da resolução de 1 metro o de verificar a utilização das metodologias propostas para varreduras LASER com densidade média maior, o que nos sistemas atuais pode ser de até 20 pontos por metro quadrado (KIM *et al.*, 2009). Tendo em vista a densidade média da varredura LASER original, salienta-se que para as células com resolução de 1 metro a quantidade de pontos armazenados foi próxima de 2 pontos por célula.

Em cada grade foram armazenados e codificados os 6.490.960 pontos LASER. A codificação empregada permite determinar se o ponto foi mensurado pelo primeiro, último ou ambos os retornos. As dimensões das grade regulares são:

- grade regular de 10 metros: 250 linhas por 210 colunas;
- grade regular de 5 metros: 500 linhas por 420 colunas;
- grade regular de 1 metro: 2500 linhas por 2100 colunas.

A Figura 21 mostra uma célula da grade regular de 10 metros. Nesta célula foram alocados 113 pontos referentes ao primeiro e último retorno, cujas coordenadas planimétricas estão contidas nos limites da célula apresentada.

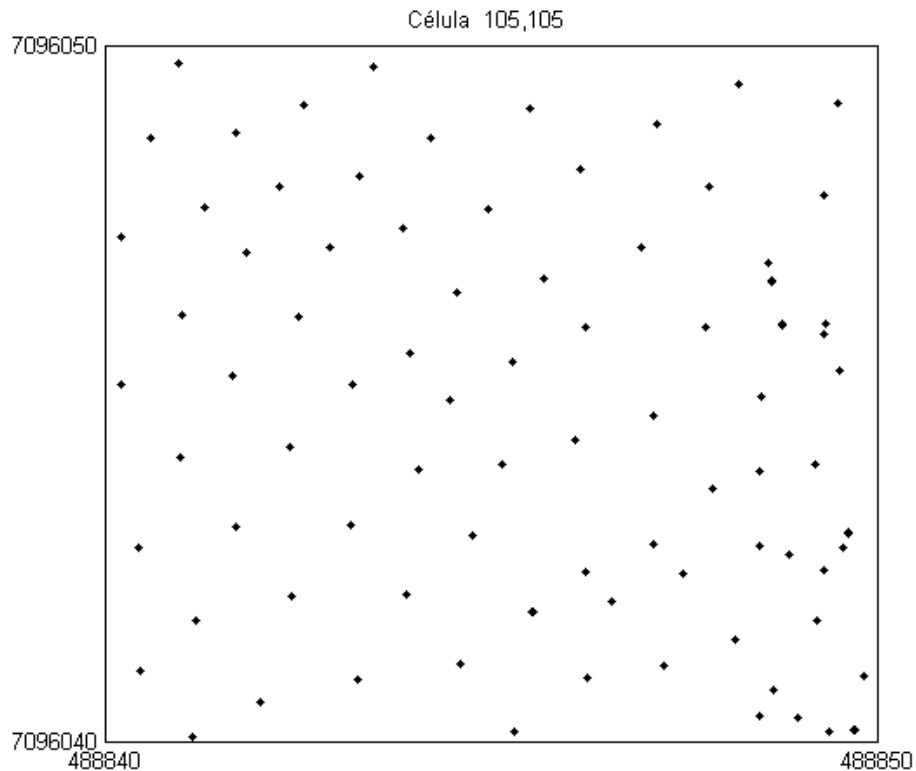


FIGURA 21 - EXEMPLO DE UMA CÉLULA DA GRADE DE 10 METROS

Uma vantagem do método de armazenagem proposto é a redução substancial no tempo de processamento dos dados brutos. Os testes iniciais com os dados armazenados no formato original consumiram 20 vezes mais tempo de processamento quando comparados com a armazenagem na forma de grade regular.

### 3.3.5 Determinação do Modelo Digital do Terreno

A altura da cobertura vegetal foi utilizada como um dos parâmetros na classificação da grade LASER de acordo com as formações vegetais detectadas pela fotointerpretação e reambulação. Para a determinação deste parâmetro foi

necessário a implementação de um algoritmo para separação de pontos no solo e acima deste.

O algoritmo implementado baseou-se na metodologia proposta por Zaksek e Pfeifer (2006). Estes autores adaptaram um filtro morfológico visando a sua utilização em áreas cobertas por florestas e com topografia íngreme.

Para melhorar o filtro morfológico o plano horizontal, normalmente empregado, foi substituído por uma superfície de tendência local, isto é, um plano ajustado para cada célula da grade de 10 metros. A partir da superfície local pode-se aplicar o filtro para a separação dos pontos que independe da direção (isotropia).

Resumidamente, o filtro compara uma diferença de elevação entre dois pontos com a máxima diferença de elevação permitida, relacionando ainda a distância horizontal entre os dois pontos. O núcleo do filtro é a superfície de um cone obtida através da minimização da classificação dos erros, que foi o método empregado na pesquisa.

A separação dos pontos localizados na superfície topográfica e acima desta, foi implementada para a grade regular com resolução de 10 metros. A resolução desta grade tomando por referência os testes realizados por Zaksek e Pfeifer (2006), bem como, o fraco desempenho do filtro quando testadas resoluções diferentes para a grade.

O processo de filtragem proposto é iterativo e neste estudo foram necessários três iterações até que o critério de parada fosse alcançado para todas as células da grade. Como critério de parada utilizou-se a análise dos resíduos do ajustamento de observações para a determinação do plano de tendência local. Definiu-se como máximo resíduo para este ajuste 0,5 metro, valor sugerido por Zaksek e Pfeifer (2006).

A seguir são apresentados os passos necessários a filtragem dos pontos LASER armazenados na grade regular:

1) Definição de uma superfície de tendência local: na primeira iteração empregaram-se somente os pontos referentes ao primeiro retorno para a geração da superfície de tendência local. A idéia do uso de dados do primeiro retorno para selecionar pontos no solo é inédita (ZAKSEK;PFEIFER, 2006). De acordo com estes autores, estes dados predizem a morfologia do terreno abaixo da vegetação, pois a elevação das copas não muda rapidamente dentro da célula. Pode-se então supor que a estrutura florestal é homogênea dentro de uma célula o que conduz para a

determinação da tendência da vegetação e da superfície topográfica. O cálculo da superfície de tendência para cada célula é obtido através da equação do plano, cujo formato é:

$$z_i - a \cdot x_i - b \cdot y_i - d = 0 \quad (102)$$

Sendo:

- $x_i, y_i$  as coordenadas planas de um ponto  $i$  e
- $a, b, d$  os coeficientes definidores do plano local.

No formato utilizado para a equação (102) o ajustamento do plano em cada célula da grade, é obtido através do método combinado:

$$F(L_a, X_a) = 0 \quad (103)$$

O ajustamento combinado foi processado individualmente para cada célula da grade, sendo o número mínimo de pontos para a determinação dos planos locais com ajustamento em cada célula igual a 4. Maiores detalhes sobre o método de ajustamento combinado podem ser encontrados em Gemael (1994).

2ª) Cálculo da inclinação do cone: depois que um a superfície de tendência local foi ajustada para cada célula, procedeu-se a obtenção da superfície que separou os pontos, que é um cone, calculado para cada célula da grade.

A inclinação do cone foi determinada a partir da análise da inclinação das células vizinhas. Para o cálculo da inclinação utilizou-se uma solução comumente empregada em programas GIS (*Geographic Information System*) comerciais (ESRI, 2005):

$$C = D + E \quad (104)$$

Onde  $C$  é a inclinação padrão do cone e:

$$D = \frac{\frac{h_2 + h_3}{2} - h_0}{L^2} \quad e \quad E = \frac{\frac{h_1 + h_4}{2} - h_0}{L^2} \quad (105)$$

Sendo que  $h_i$  ( $i = 0, 1, 2, 3$  e  $4$ ) são as altitudes médias para as células e  $L$  é a resolução da grade retangular, ou seja, 10 metros. A distribuição das células vizinhas empregadas no cálculo da inclinação do cone é mostrada na Figura 22.

	$h_1$	
$h_2$	$h_0$	$h_4$
	$h_3$	

FIGURA 22 –CÁLCULO DA INCLINAÇÃO DO CONE.  
FONTE: Adaptado de Zaksek e Pfeifer, 2006.

Lembrando que para a primeira iteração foram utilizados para o cálculo da inclinação somente os pontos tridimensionais determinados pelo primeiro retorno LASER.

3ª) Filtragem dos pontos LASER: neste passo todos os pontos alocados nas células (primeiro e último retorno) foram classificados através da seguinte equação :

$$H_j^{min}(dist) = a \cdot x_j + b \cdot y_j + d + v_j - C \cdot dist - f(\alpha) \cdot \sqrt{2} \cdot \sigma \quad (106)$$

Onde:

- a,b,d: parâmetros obtidos no ajustamento do plano local para cada célula;
- C: inclinação da célula;
- dist = resolução da grade LASER, isto é, 10 metros;
- $f(\alpha) = 95\%$  conforme sugerido por ZAKSEK & PFEIFER (2006);
- $\sigma = 0,50$  metro. Refere-se à precisão do sistema LASER scanner.

O valor de  $v$  foi obtido através da diferença entre a altitude registrada pelo sistema LIDAR e a altitude calculada através da equação do plano ajustado para a célula.

O valor de  $H$  resultou positivo quando o ponto analisado situou-se acima do cone e negativo quando se situou abaixo do cone.

O ponto foi classificado como pertencente à superfície topográfica se está abaixo do cone.

Nas iterações seguintes o primeiro e segundo passos são processados novamente, utilizando-se os pontos classificados como pertencentes a superfície topográfica na iteração anterior.

O critério para o término do processo de filtragem recai sobre o primeiro passo, ou seja, o cálculo da superfície de tendência local. O processo de iteração termina quando os resíduos do ajustamento do plano estão no intervalo de -0,5 a 0,5 metro.

4) Verificação da acuracidade da Filtragem: para verificar a acurácia do filtro morfológico adaptado foram empregadas como referência as observações obtidas pelos perfis fotogramétricos.

Para possibilitar a comparação, os pontos fotogramétricos também foram armazenados em uma grade regular com resolução de 10 metros. Para cada célula da grade fotogramétrica foram ajustados os planos locais e então determinada a altitude para o centro de cada célula da grade.

O mesmo procedimento foi implementado para a grade contendo os pontos LASER, porém, o ajuste foi obtido empregando-se os pontos filtrados (classificados) como pertencentes a superfície topográfica.

As duas grades foram sobrepostas, e então, calculadas as diferenças de altitude entre as células das duas grades, gerando-se com isso valores positivos e negativos. Estes valores relativos a diferença de altitude entre as células da grade LASER e da grade fotogramétrica foram armazenados separadamente em uma matriz. Estes dados possibilitaram a verificação das discrepâncias entre o MDT derivado a partir dos dados obtidos pelo processo de filtragem matemática e pela fotogrametria.

Outro critério empregado para verificar a acuidade do processo de filtragem de pontos foi a interpolação das curvas de nível. Foram criadas as curvas de nível a partir dos dados provenientes da filtragem matemática e também da fotogrametria. As curvas foram sobrepostas e procedida a análise visual.

### 3.3.6 Determinação da Altura da Cobertura Vegetal

A altura da cobertura vegetal foi empregada nas atividades de reambulação como um dos parâmetros auxiliares na classificação do estágio sucessional das formações vegetais naturais.

Para calcular esta variável a partir dos pontos LASER armazenados na grade regular, inicialmente, definiu-se a superfície topográfica para a área dos experimentos empregando-se a filtragem matemática.

Os pontos filtrados possibilitaram o ajuste de planos locais para cada uma das células da grade regular.

Na seqüência, processando-se individualmente cada célula determinou-se a altura da cobertura vegetal dos pontos armazenados. As coordenadas bi-dimensionais de cada um dos pontos foram aplicadas a equação do plano local. A altura da cobertura vegetal é obtida através de:

$$\Delta h = h_m - h_c \quad (107)$$

Onde:

- $\Delta h$  : altura da cobertura vegetal;
- $h_m$ : altitude do ponto determinada pelo sistema LIDAR;
- $h_c$  : altitude calculada a partir da equação (102).

A Figura 23 apresenta uma célula da grade regular com um plano local ajustado (destacado pelos limites em tom vermelho). Este plano foi ajustado a partir dos pontos selecionados como pertencentes a superfície topográfica. A altura da cobertura vegetal ( $\Delta h$ ) é a distância vertical do ponto *laser* até o plano ajustado.

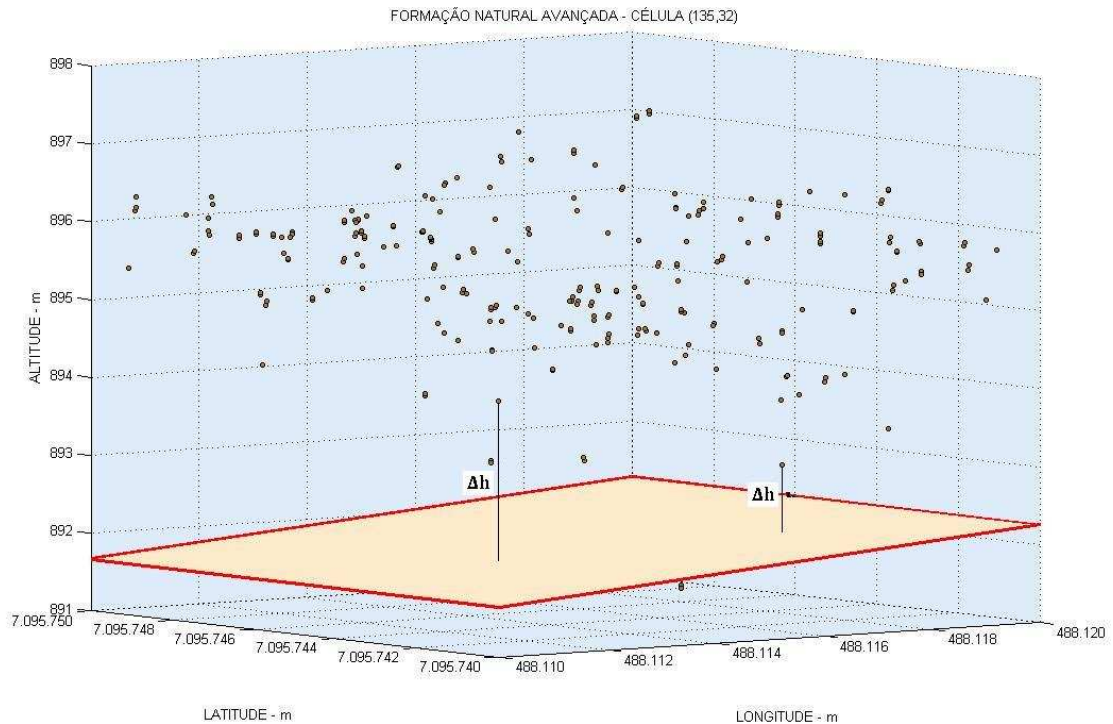


FIGURA 23 – VISUALIZAÇÃO DA VARIÁVEL ALTURA DA COBERTURA VEGETAL PARA UMA CÉLULA OCUPADA POR FLORESTA

Ao final deste processo cada um dos pontos LASER apresentava as seguintes variáveis: coordenadas espaciais Leste (E), Norte (N), altitude ortométrica (h), intensidade de refletância (I) e altura da cobertura vegetal ( $\Delta h$ ).

Finalmente, os pontos armazenados na grade regular de 10 metros, também foram armazenados nas grades com resolução de 5 metros e de 1 metro.

### 3.3.7 Obtenção das Amostras Referentes à Cobertura Vegetal

Neste ponto, cabe lembrar que o objetivo final desta pesquisa foi a classificação das células das grades regulares nas quais está armazenada a nuvem de pontos LASER. A classificação compreende a separação das células em classes de acordo com as coberturas vegetais listadas no quadro 02.

Foram testados como classificadores a Função Discriminante de Fisher, a Função Discriminante Quadrática e a Regressão Logística com Camada Oculta.

Como cofatores destas funções foram utilizados:

- h: altitude ortométrica média;



- $\sigma_h$  : desvio padrão das altitudes ortométricas;
- $I$ : intensidade média;
- $\sigma_I$  : desvio padrão das intensidades;
- $\Delta h$ : altura média da cobertura vegetal;
- $D_e$  : densidade de pontos.

Os valores dos cofatores foram calculados para cada uma das células das grades regulares.

Foram observados os seguintes critérios para a escolha dos cofatores:

- A altitude média está relacionada com a ocorrência de determinadas espécies florestais que, por exemplo, tem preferência por locais com solo mais seco, ou seja, áreas mais elevadas e com melhor condição de drenagem. Os locais com menores altitudes na área de estudo são ocupados com áreas alagáveis e com formações vegetais peculiares.

- A intensidade média é o valor da refletância da superfície imageada pelo feixe LASER e, portanto, referindo-se as características espectrais do alvo. Esta variável é empregada na maioria das pesquisas relacionadas com a classificação de dados LASER, citando como exemplo, o trabalho desenvolvido por KIM *et al.* (2009).

- A altura média da cobertura vegetal é complementar aos cofatores altitude e intensidade média que são similares para os locais cobertos por formações naturais com regeneração avançadas e médias. Ademais, O estágio da formação natural está diretamente relacionado à altura, quanto mais avançado o estágio de regeneração maior a altura da cobertura vegetal.

- O desvio padrão da altitude e da intensidade tem relação direta com a textura do alvo. Os autores PFEIFER *et al.* (2004) propuseram o cálculo da textura a partir dos pontos LASER e sua utilização na classificação da cobertura do solo.

- A densidade de pontos é um valor que traduz a quantidade de pontos por metro quadrado. Ao observar uma área coberta por florestas é possível verificar mais de uma reflexão do pulso LASER que ocorre nas copas e abaixo destas. Em áreas ocupadas por campo ocorrerá uma única reflexão, devido a uniformidade da superfície. Na água a energia do pulso é totalmente absorvida e portanto, a densidade será muito pequena ou mesmo nula.

A fim de verificar as características apresentadas foram analisadas as células amostradas. Por exemplo, seja a célula {224,108} da grade regular de 10

metros de resolução. Esta célula corresponde a uma área ocupada por agropecuária. Ela apresenta as seguintes valores:

- $h = 834,800$  m;
- $l = 87,3$ ;
- $\sigma_h = 0,3666$  m;
- $\sigma_l = 9,6796$ ;
- $De = 0,64$  ponto/ $m^2$ ;
- $\Delta h = 0,35$  m.

Considerando agora a célula {190,51} que é uma área coberta por formação natural avançada os valores são os seguintes:

- $h = 975,845$  m;
- $l = 142,5$ ;
- $\sigma_h = 3,2237$  m;
- $\sigma_l = 38,1533$ ;
- $De = 1,44$  ponto/ $m^2$ ;
- $\Delta h = 14,57$  m.

A classificação da cobertura vegetal inicia com o ajuste das funções de discriminação. Este ajuste ocorre a partir de amostras georreferenciadas referentes às classes almeçadas.

O processo de amostragem compreende a coleta dos valores para os cofatores cujas classes das células da grade sejam conhecidas. Neste intuito, foram delimitadas áreas amostrais para as classes temáticas tomando por referência a base cartográfica compilada através da restituição fotogramétrica.

A obtenção das amostras observou os critérios listados por Centeno (2004), segundo o qual, a amostragem é o primeiro passo para uma boa classificação supervisionada, pois nela as características das classes são definidas, características que permanecem constantes durante todo o processo. A amostragem é feita identificando áreas de cobertura conhecida e delimitando-se espacialmente estas regiões. De forma a garantir a representatividade das amostras as regiões escolhidas devem estar localizadas em locais onde as classes apareçam puras. Ademais, as áreas de treinamento devem estar distribuídas ao longo de toda a região a ser classificada, procurando cobrir todas as diferentes situações nas quais

a classe aparece. Outro critério é o tamanho das amostras, sendo desejados conjuntos maiores que 50 células.

Para cada formação vegetal selecionaram-se 2 áreas amostrais sendo delimitadas 24 áreas a partir das quais foram obtidas as amostras nas células das grades. A Figura 24 apresenta a distribuição das áreas amostrais para a área de estudo.

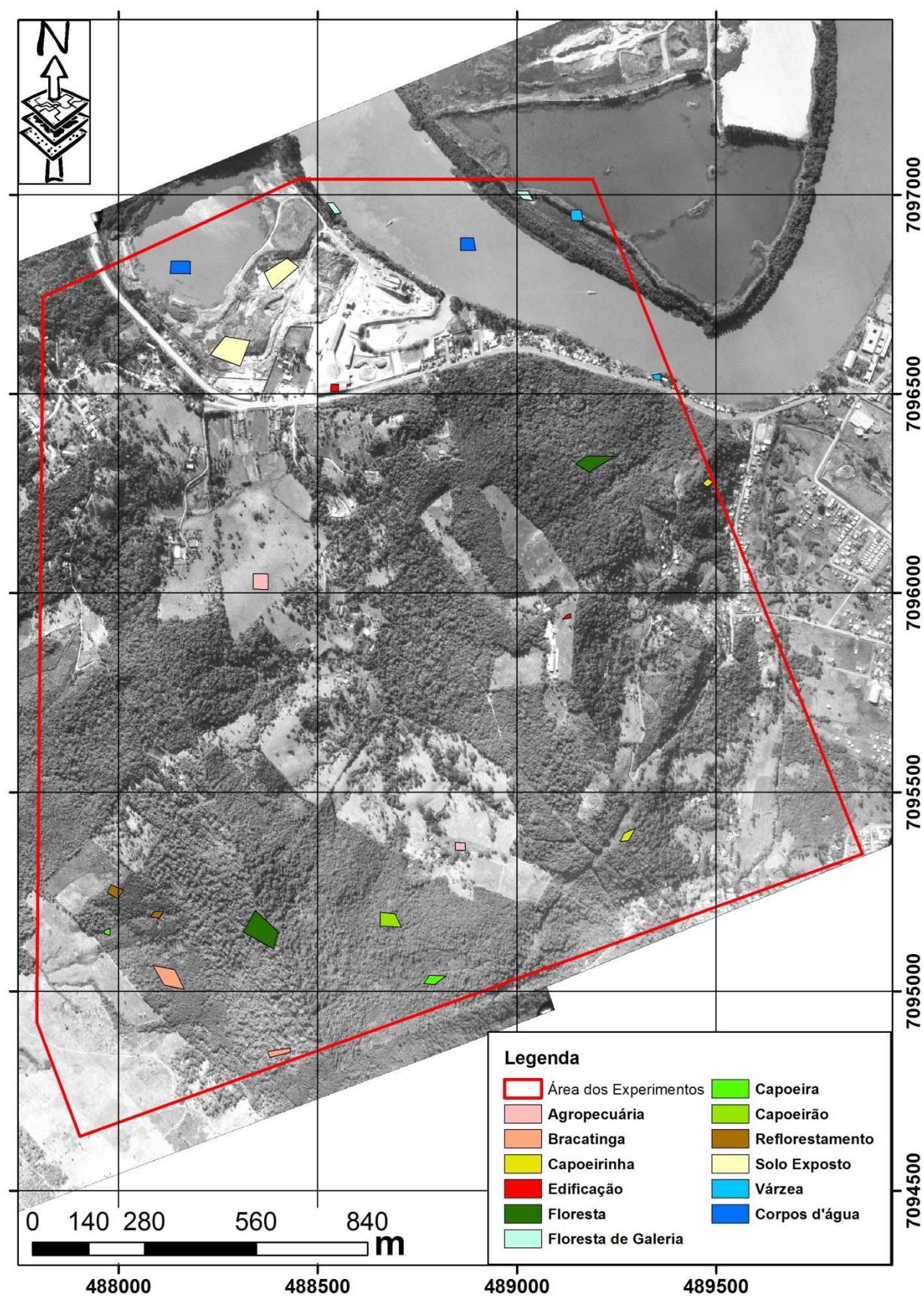


FIGURA 24 – DELIMITAÇÃO DAS ÁREAS AMOSTRAIS

Os valores obtidos no processo de amostragem foram armazenados no formato matricial. Como foram criadas três grades regulares (10 m, 5 m e 1 m) obtiveram-se três conjuntos de amostras e, portanto, três matrizes amostrais (X). O formato destas matrizes:

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} & x_{15} & x_{16} & x_{17} \\ x_{21} & x_{22} & x_{23} & x_{24} & x_{25} & x_{26} & x_{27} \\ \vdots & & & & & & \\ x_{n1} & x_{n2} & x_{n3} & x_{n4} & x_{n5} & x_{n6} & x_{n7} \end{bmatrix}$$

Sendo  $n$  o número total de células amostradas.

Foram armazenados da primeira para a sétima coluna os seguintes valores: tipo de cobertura vegetal, altitude média, intensidade média, altura média da cobertura vegetal, desvio padrão da altitude, desvio padrão da intensidade e densidade de pontos na célula.

### 3.3.8 Ajuste das Funções Discriminantes

O ajuste tem por finalidade a obtenção das funções de discriminação estimadas a partir do processamento das amostras obtidas. Estas funções de discriminação estimadas possibilitaram a classificação das células que compõe as três grades regulares.

A qualidade da função discriminante não só depende da identificação correta da distribuição de probabilidades do vetor resposta em cada classe de interesse mas também das variáveis escolhidas para a sua elaboração. Quando se considera a utilização da análise discriminante, é essencial que se tenha medido, nos elementos amostrais, variáveis que possam realmente distinguir as populações, caso contrário a qualidade de ajuste da regra de discriminação estará comprometida (MINGOTI,2005).

Procedeu-se o teste F de Snedecor entre as 12 classes temáticas consideradas em relação a cada variável individualmente. Aquelas variáveis com valores da estatística F mais significativos estão relacionadas às variáveis mais importante da discriminação.

A estatística F determina a relação significativa entre a variável dependente e o conjunto de variáveis explicativas, testando-se as hipóteses:  $H_0$  (hipótese nula) se  $\beta_1=0$  (não existe relação linear entre a variável dependente e a variável explicativa), e  $H_a$  (hipótese alternativa) se  $\beta_1 \neq 0$ , mostrando a existência de relação linear entre as variáveis dependente e independente. Se o valor de F estatístico for maior que o F crítico, rejeita-se  $H_0$ .

Enfatizando a importância da distribuição F, pode-se dizer que esta é sempre usada para verificar se duas amostras pertencem a mesma população, como também para comparar duas ou mais médias simultaneamente, denominada como análise de variância, sempre que populações tenham distribuição normal. Esta distribuição é sempre contínua e positiva e seus valores pertencem ao intervalo  $(0, +\infty)$ , com dois parâmetros com graus de liberdade (numerador e denominador).

O teste F foi implementado através do método *stepwise forward*. De acordo com Mingoti (2005), no primeiro passo faz-se o teste F para cada uma das  $p$  variáveis candidatas a entrar no modelo separadamente. Dentre as variáveis significativas, dá entrada no modelo de discriminação aquela de maior significância, ou de maior poder de discriminação dos grupos. Se não houver qualquer variável significativa, o modelo é interrompido. No segundo passo, os F parciais de entrada são calculados levando-se em consideração apenas as variáveis que já entraram e permaneceram no modelo. Os valores calculados de F parciais, para as variáveis candidatas são comparados com um valor de F de referência e que é escolhido a priori.

O teste proposto mostrou a significância estatísticas de todos os cofatores elencados.

Para o ajuste das funções de discriminação estimadas foram selecionadas as seguintes combinações de variáveis:

- Altitude média e intensidade média
- Altitude média, intensidade média e altura média da cobertura vegetal e
- Altitude média, intensidade média, altura média da cobertura vegetal, desvio padrão da altitude, desvio padrão da intensidade e densidade de pontos na célula.

A seguir são apresentadas as etapas de trabalho para o ajuste das funções de discriminação considerando-se a terceira combinação de variáveis apresentada.

### 3.3.8.1 Ajuste da Função Discriminante de Fisher

Os passos necessários ao ajuste da Função Discriminante empregando o modelo linear de Fisher são:

1º) Cálculo dos vetores de médias amostrais: foram determinadas as médias amostrais de cada variável para cada uma das 12 classes de cobertura vegetal. O cálculo das médias amostrais é obtido por:

$$\bar{x}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \quad (108)$$

Sendo que:

$i$  = tamanho da amostra para a classe e

$j = 1, 2, \dots, 6$ .

As dimensões da matriz de médias amostrais, que contém os vetores é de  $n$  linhas por 6 colunas, sendo  $n$  o tamanho total da amostra.

2º) Cálculo do vetor de média global: determinou-se os valores médios amostrais de cada variável considerando-se todas as classes temáticas. O vetor médio global é calculado por:

$$\bar{x} = \frac{1}{\sum_{i=1}^g n_i} \sum_{i=1}^g n_i \bar{x}_i \quad (109)$$

Na equação acima  $g$  refere-se ao número de classes, ou seja, 12.

3º) Cálculo das matrizes de co-variância amostral ( $S_i$ ), com  $i = 1, 2, \dots, g$ .

4º) Cálculo da matriz de co-variância entre grupos ( $S_p$ ), obtida através da equação:

$$S_p = \left( \sum_{i=1}^g (n_i - 1) S_i \right) \frac{1}{(\sum_{i=1}^g n_i) - g} \quad (110)$$

$S_p$  têm dimensões de 6 linhas por 6 colunas.

5º) Cálculo da matriz de variação entre os grupos ( $\hat{B}_0$ ): o cálculo é obtido através da equação:

$$\hat{B}_0 = \sum_{i=1}^g (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' \quad (111)$$

A matriz  $\hat{B}_0$  apresenta dimensões de 6 linhas por 6 colunas.

6º) Cálculo do estimador de co-variância  $W$ : este estimador é obtido através da equação:

$$W = \left[ \left( \sum_{i=1}^g n_i \right) - g \right] S_p \quad (112)$$

A matriz  $W$  apresenta dimensões de 6 linhas por seis colunas.

7º) Cálculo da expressão:  $W^{-1}\hat{B}_0$  com dimensões 6 linhas por 6 colunas;

8º) Cálculos dos autovalores ( $\lambda$ ) e autovetores ( $e$ ) da matriz obtida no 7º passo;

9º) Determinação do espaço canônico, isto é, da dimensionalidade das funções discriminantes. O seu cálculo é obtido por:

$$n = \min \{g - 1, p\} \quad (113)$$

Sendo  $n$  a dimensão do espaço canônico,  $g$  o número de classes temáticas e  $p$  o número de cofatores. No caso, a dimensão do espaço canônico resulta igual a 6.



10) Obtenção das funções lineares discriminantes aproximadas: A primeira função discriminante é obtida através do primeiro autovetor  $\hat{e}_{11}$  e  $\hat{e}_{21}$  de  $W^{-1}B$ . A segunda função discriminante será obtida pelo segundo autovetor  $\hat{e}_{12}$  e  $\hat{e}_{22}$  de  $W^{-1}B$  e assim até a 6ª equação de discriminação.

11) Cálculos dos centróides das classes: Nesta etapa são determinados os centróides para as classes temáticas. Os valores médios amostrais de cada classe temática são aplicados nas funções de discriminação, que são os 12 centróides com dimensão  $n$ .

Após o término do ajuste das seis funções de discriminação aproximada obtidas a partir das amostras da grade regular de 10 metros, obteve-se ajuste destas funções para as amostras provenientes das grades regulares de 5 metros e 1 metro respectivamente.

Obteve-se assim um conjunto de funções discriminantes para a grade de 10 metros, 5 metros e 1 metro com a combinação de variáveis selecionadas.

### 3.3.8.2 Ajuste da Função Discriminante Quadrática

A motivação para utilização da Função Discriminante Quadrática é que esse modelo considera as matrizes de covariâncias das classes diferentes além de possibilitar a separação não linear entre as classes.

Os passos necessários ao ajuste das funções discriminantes quadráticas são os seguintes:

- 1) Entrada da matriz amostral  $X$  referente as classes temáticas;
- 2) Definição da probabilidade a priori a ser empregada na estimação da função discriminante: optou-se por trabalhar com probabilidades iguais. A outra alternativa é utilizar probabilidade proporcional aos tamanhos das amostras;
- 3) Cálculo da média das classes temáticas;
- 4) Cálculo da matriz de co-variâncias das classes temáticas;
- 5) Cálculo das constantes das funções discriminantes;
- 6) Cálculo dos coeficientes de primeiro grau das funções discriminantes: o resultado é um vetor de coeficientes para cada classe;

7) Cálculo dos coeficientes de segundo grau das funções discriminantes: a saída é um vetor de coeficientes para cada classe;

8) Montagem das 12 equações de discriminação de acordo com a equação 54.

### 3.3.8.3 Ajuste da função discriminante Logística Polinômica com Camada Oculta

A Regressão Logística Polinômica com Camada Oculta (RLO) é a adaptação do processo mostrado por Hosmer e Lemeshow (1989) e neste trabalho segue a seqüência mostrada por estes autores.

Os passos para o ajuste das 11 funções de verossimilhança foram:

1) Entrada da matriz amostral X referente as classes temáticas;

2) Definição da classe temática de referência: nos testes realizados a classe temática de referência selecionada é a primeira alocada na matriz amostral.

3) Cálculo equações de verossimilhança (Y): este passo é obtido através da resolução das 11 equações de verossimilhança obtidas a partir da equação (79). Observando o citado por Hosmer e Lemeshow (1989), para o cálculo dos estimadores empregou-se o método de Newton Raphson, definido como:

Seja  $f$  uma função diferenciável em um intervalo aberto contendo um ponto  $c$  tal que  $f(c) = 0$ . As etapas a seguir permitem aproximar o valor de  $c$ :

a) Escolha de uma estimativa inicial  $x_1$  próxima de  $c$ ;

b) Determinação de uma nova aproximação:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} \quad (114)$$

Sendo  $n$  o número da iteração e  $f'(x_n)$  a derivada parcial da função logística  $f(x_n)$ .

c) Se  $|x_n - x_{n+1}|$  é menor do que 0,0001, toma-se  $x_{n+1}$  como a aproximação desejada. Caso contrário, repete-se o passo b e obtem-se uma nova estimativa; é que ele não converge quando há separação total de pelo menos um grupo em relação aos demais. Também vale lembrar que neste caso serão 11 funções discriminantes, pois cada grupo será comparado com o grupo de referência.

4) Contaminação do vetor resposta  $Y$ : neste passo calculou-se os valores dos deltas ( $\delta_s$  e  $\delta_f$ ). Os deltas são os valores que possibilitam o cálculo de  $\hat{Y}$  e que propiciam a camada oculta ao método de regressão logístico múltiplo com resposta politômica. Como são 12 grupos, determina-se o mesmo número de valores para  $\delta$ . A notação adotada nesta pesquisa é a mesma utilizada por Guimarães (2006).

Fazendo  $\delta_{jk}$  com  $j, k = 1, 2, \dots, g$  e  $\delta_{jj} = \delta$  e  $\delta_{jk} = \frac{1-\delta}{g-1}$ , basta escolher um valor de  $\delta$ . Uma boa escolha é  $\delta = 0,99$ . O objetivo desta escolha é trocar a variável  $Y$  por  $\hat{Y}$ , de tal maneira que a resposta não seja exatamente 0 ou 1, mas um valor entre 0 e 1. Os valores de  $\hat{Y}$  são calculados por:

$$\hat{y}_{ij} = \sum y_{ik} \delta_{kj} \quad (115)$$

Por exemplo, o vetor de respostas para a área dos experimentos apresenta 12 classes. Se uma determinada observação pertence a quarta classe temática, isto é,  $Y = 4$ , o vetor fica (0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0). Então, para cada  $Y_i$ ,  $i = 1, \dots, n$ , tem-se um vetor  $y_{ij}$ ,  $j = 1, \dots, s$ . Deste modo  $y_{ij} = 1$  se a observação pertence a classe  $j$ , e  $y_{ij} = 0$  se a observação não pertence a classe  $j$ .

Neste caso:

$$\delta = 0,99,$$

$$\delta_{jj} = 0,99 \text{ e}$$

$$\delta_{jk} = (1 - 0,99)/(12 - 1) = 0.001$$

Fazendo a substituição, o vetor  $Y = (0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0)$  resulta em  $\hat{Y} = (0.001, 0.001, 0.001, 0.99, 0.001, 0.001, 0.001, 0.001, 0.001, 0.001, 0.001, 0.001)$ .

O modelo de regressão logístico com esta transformação passa de côncavo para estritamente côncavo.

5) Repete-se o terceiro passo, porém, empregando -se o vetor calculado  $\hat{Y}$ .

Ao final deste processo foram obtidos 3 conjuntos de 11 equações de verossimilhança referentes a cada uma das três grades regulares.

A próxima etapa consistirá na classificação das células das grades regulares.

### 3.3.9 Determinação da Cobertura Vegetal Através da Classificação das Grades Regulares

Com o treinamento das funções de discriminação concluído foi possível então classificar os pontos LASER armazenados nas células de cada uma das três grades regulares.

Cada uma das células das grades regulares foi classificada individualmente em uma das 12 classes temáticas referentes a cobertura vegetal. Detalha-se a seguir como foi procedida a classificação.

#### 3.3.9.1 Classificação Baseada na Função Discriminante de Fisher

Considerando-se a classificação de uma célula  $(i,j)$  da grade regular, os passos necessários a classificação as Funções Discriminantes Aproximadas de Fisher são:

1) Cálculo dos cofatores para a célula: são calculados os cofatores referentes aos pontos armazenados na célula;

2) Cálculo das funções de discriminação: os valores obtidos no primeiro passo são aplicados nas funções de discriminação e obtêm-se as coordenadas canônicas para a célula analisada.;

3ª) Cálculo das distâncias euclidianas: as coordenadas canônicas da célula são utilizadas para o cálculo das distâncias euclidianas em relação aos 12 centróides de cada uma das classes temáticas;

4ª) Classificação da célula: a menor das 12 distâncias euclidianas indicará a qual classe pertence a célula que está sendo analisada.

5ª) Armazenamento da classificação: o resultado da classificação obtida para a célula é um número inteiro variando de 1 até 12. Este número relativo a classificação foi armazenado na posição  $(i,j)$  de uma matriz, com dimensões idênticas a grade regular.

### 3.3.9.2 Classificação Baseada na Função Discriminante Quadrática

Os passos a seguir apresentados possibilitaram a classificação de uma célula  $(i,j)$  empregando-se a Função Discriminante Quadrática:

1ª) Cálculo das variáveis para a célula analisada: são calculadas os cofatores relativos aos pontos armazenados na célula que está sendo classificada;

2ª) Aplicação dos valores dos cofatores calculados em cada uma das 12 funções de discriminação;

3ª) Determinação das probabilidades: cada uma das 12 funções de discriminação retornará um valor de probabilidade, variando de 0 a 1;

4ª) O maior dos 12 valores de probabilidade calculados indicará a qual classe temática pertence a célula  $(i,j)$  que está sendo analisada;

5ª) Armazenamento da classificação: o resultado da classificação obtida para a célula  $(i,j)$  é então armazenado na posição  $(i,j)$  de uma matriz, com dimensões idênticas a grade regular.

### 3.3.9.3 Classificação Baseada na RLO

Os procedimentos para a classificação das células empregando a RLO foram:

- 1) Cálculo dos cofatores para a célula analisada: são calculadas as variáveis relativas aos pontos armazenados na célula que está sendo classificada;
- 2) Cálculos das equações de verossimilhança ( $\hat{Y}$ );
- 3) Cálculo das probabilidades condicionadas empregando-se a equação (76);
- 4) Determinação da classe temática: o maior dos valores de probabilidade calculados indica a qual classe temática pertence a célula analisada;
- 5) Armazenamento da classificação: o resultado da classificação obtida para a célula é armazenado em uma matriz, com dimensões idênticas a grade regular.

Concluída a classificação das células das grades regulares procedeu-se a análise dos resultados das classificações.

#### 3.3.9.4 Determinação e Análise dos Índices de Acerto das Funções Discriminantes

Ao final do processo de classificação foram obtidas matrizes cujos elementos contêm a classificação temática para cada célula armazenadas nas grades regulares.

Elaborou-se então uma metodologia para a determinação dos índices de acerto de cada uma das funções discriminantes. As principais etapas de trabalho são apresentadas:

- 1) Compatibilidade das coordenadas: antes de realizar comparações entre as grades classificadas e a base cartográfica compilada pela fotointerpretação, verificou-se a compatibilidade das coordenadas. Neste intuito, utilizou-se a imagem de intensidade derivada das observações LASER e a base cartográfica proveniente da restituição fotogramétrica para a escolha de pontos bem definidos nestas duas fontes de dados. Foram selecionados pontos relativos a intersecção de vias e bordas de áreas ocupadas por agropecuária;

- 2) Obtenção da base cartográfica em formato matricial: para tornar possível a análise dos índices de acerto das classificações das grades, a base cartográfica

especializada pela fotointerpretação foi considerada como referência. Este mapa em formato vetorial foi convertido para o formato matricial empregando uma função disponibilizada pelo programa Arcgis 9.2. Ao final desta tarefa, obteve-se três matrizes, cujas resoluções espaciais são de 10 metros, 5 metros e 1 metro. Em cada célula destas matrizes está armazenado um número inteiro, variando de 1 até 12, que se refere a qual classe de cobertura vegetal associada a célula;

3) Sobreposição da base cartográfica em formato matricial com as matrizes classificadas pelas funções discriminantes: nesta etapa as matrizes referentes a verdade são sobrepostas as matrizes geradas pelas funções discriminantes;

4) Montagem da matriz de confusão: a matriz de confusão é uma forma de representar o desempenho do processo de classificação. Nesta matriz são confrontadas duas situações: as linhas correspondem à classe correta e as colunas à classe estimada no processo de classificação. Cada célula da matriz de confusão armazena a quantidade de elementos que deveriam ser classificados como a classe especificada na linha e que acabaram sendo classificados segundo a classe especificada na coluna. Em uma situação ideal, onde todos os elementos são classificados corretamente, a matriz apresentará valores apenas na diagonal (CENTENO,2004). Para obter a matriz de confusão as células das matrizes classificadas são comparadas com as células da matriz que contém a verdade. São totalizadas como acerto as células da matriz classificada cujos valores coincidirem com a matriz que contém a verdade.

No resultado da classificação distinguem-se dois tipos principais de erro: erros de omissão e erros de inclusão. Os erros de omissão referem-se as células pertencentes a uma classe que foram classificadas erroneamente como sendo de outra, isto é, foram omitidas da classe (CENTENO,2004). A determinação deste erro é obtida através da equação:

$$E_{om(i)} = \frac{SL(i) - L(i, i)}{SL(i)} \quad (116)$$

Da equação (116):

$SL(i)$  - soma parcial da linha  $i$ ;

$L(i, i)$  - elemento da diagonal na linha  $i$ .

Já os erros de inclusão correspondem as células que, sendo na realidade de outra classe, foram incluídos na classe considerada (CENTENO, 2004). Este erro pode ser calculado através da equação:

$$E_{in(i)} = \frac{SC(i) - C(i, i)}{SC(i)} \quad (117)$$

Na equação (117):

$SC(i)$  - soma parcial da coluna  $i$ ;

$C(i, i)$  - elemento da diagonal na coluna  $i$ ;

Ainda de acordo com Centeno(2004), os erros de omissão e inclusão são elementos que podem ser empregados para avaliar o quanto uma classe foi sub ou superestimada. Os erros de omissão refletem a precisão com que uma classe foi definida. Os erros de inclusão indicam a superestimação de uma classe.

Os erros de omissão, que são avaliados na linha da matriz de confusão, estão associados à exatidão relacionada a classificação que pode ser medida pela taxa de células corretamente classificadas em cada classe. Quanto maior esta taxa, maior será o sucesso alcançado pelo classificador utilizado.

##### 5) Análise dos resultados obtidos com base nas matrizes de confusão.

A Figura 25 apresenta um fluxograma demonstrando as atividades relacionadas com a classificação das três grades regulares.



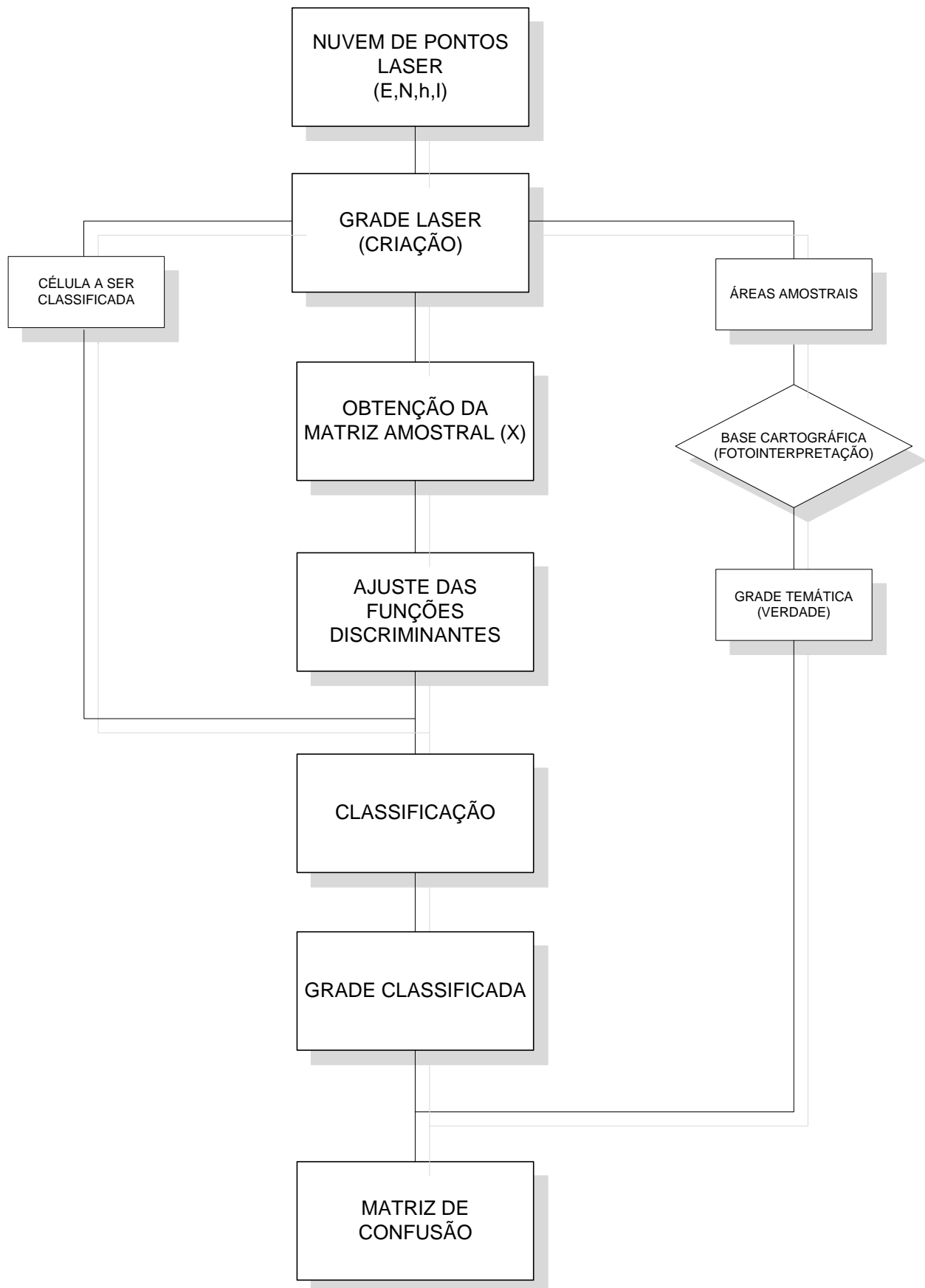


FIGURA 25 – ETAPAS DE CLASSIFICAÇÃO DA COBERTURA VEGETAL

## 4 RESULTADOS E DISCUSSÕES

### 4.1 APOIO DE CAMPO

A fim de obter posicionamento tridimensional referido a um Sistema Geodésico Local, os pontos de apoio foram obtidos com posicionamento GPS empregando o método estático. Outra peculiaridade foi a utilização de um receptor (base) posicionado sobre um ponto com coordenadas conhecidas localizado nas proximidades da área de estudo. Desta forma o comprimento da linha de base foi menor que cinco quilômetros e com isso foi possível obter soluções baseadas na fase da portadora L1. Estas características se traduzem em precisões elevadas, compatíveis com as necessidades do trabalho realizado.

Considerando-se o processo de restituição empregado, no qual a aquisição de dados é guiada através da marca estereoscópica com tamanho de 0,050 mm é de se esperar uma precisão na obtenção das medidas fotogramétricas equivalente a metade do tamanho da marca estereoscópica, ou seja, 0,025 mm. Como a escala média das aerofotos é de 1/10.000 a maior precisão a ser obtida na mensuração fotogramétrica será de 0,250 m. Desta forma, o apoio de campo foi executado visando obter coordenadas dos pontos com precisão compatível com as mensurações fotogramétricas, ou seja, melhor que 25 centímetros.

As observáveis GPS obtidas para os 15 pontos de apoio foram processadas no programa Astech Solutions 2.6. Os resultados deste processamento são apresentados na tabela 1.

TABELA 1 - RESULTADOS DO PROCESSAMENTO DO APOIO DE CAMPO

Ponto	E(m)	N(m)	h(m)	$\Delta E$ (m)	$\Delta N$ (m)	$\Delta$ (m)	$\Delta h$ (m)
10004	489600	7095884,134	761,960	0,002	0,007	0,007	0,015
20006	486396	7095867,147	762,034	0,008	0,009	0,012	0,022
20007	487254	7095310,352	755,833	0,004	0,001	0,004	0,003
20008	486902	7096309,014	752,531	0,001	0,001	0,001	0,004
20009	487486	7096474,567	781,283	0,023	0,014	0,027	0,020
20010	487670	7095781,142	813,005	0,018	0,020	0,027	0,037
20011	488105	7094800,900	939,201	0,001	0,003	0,003	0,003
20012	487852	7096638,789	792,508	0,034	0,042	0,054	0,057
20013	488315	7095647,504	811,793	0,007	0,007	0,010	0,005
20014	488792	7095327,905	823,930	0,002	0,002	0,003	0,003
20015	488453	7097037,945	750,571	0,105	0,078	0,131	0,207
20016	488809	7096130,596	791,075	0,122	0,054	0,133	0,099
20017	489160	7096603,877	763,803	0,090	0,083	0,122	0,092
20018	489497	7095176,018	763,479	0,012	0,009	0,015	0,018
20019	488910	7097045,063	753,229	0,035	0,045	0,057	0,027
$\bar{x}$				0,031	0,025	0,040	0,041
$\sigma$				0,041	0,028	0,048	0,055

O Sistema de coordenadas está referido ao Datum WGS-84 e a Projeção Cartográfica é a UTM com Meridiano Central 51°w.

Ainda da tabela 01 tem-se:

-  $\Delta E, \Delta N, \Delta H$ : precisões obtidas para as coordenadas E,N,H no processamento das observáveis GPS;

$$\Delta = \sqrt{\Delta E^2 + \Delta N^2} \quad (118)$$

-  $\bar{x}$  - média e

-  $\sigma$  - desvio padrão das observações.

A análise da Tabela 1 permite observar que a menor precisão planimétrica ( $\Delta$ ) obtida no processamento GPS ocorreu para o ponto 20016 com 0,133 m e a maior precisão foi obtida com o posicionamento do ponto 20008 com 0,001 m. Ao considerar a altimetria, a maior precisão altimétrica ( $\Delta h$ ) pode ser verificada para os pontos 20007 e 20014, com 0,003 m. A menor precisão altimétrica foi obtida com o ponto 20015, com o valor de 0,207 m.

Analisando ainda a Tabela 1 pode-se verificar que a precisão média para os pontos de apoio foi de 0,059 m e o desvio padrão resultou em 0,073 m.

Esta análise permitiu afirmar que os pontos de apoio têm precisão compatível com os processos de orientação absoluta necessários à restituição fotogramétrica.

## 4.2 RESTITUIÇÃO FOTOGRAMÉTRICA E REAMBULAÇÃO

A restituição fotogramétrica e a reambulação têm seus resultados divididos em três partes:

- Restituição das classes temáticas identificadas pela reambulação e fotointerpretação manual;
- Determinação dos perfis longitudinais sobre o terreno;
- Verificação da Compatibilidade entre as coordenadas do mapa temático produzido pela restituição e os dados LASER.

### 4.2.1 Restituição das Classes Temáticas

O resultado da espacialização da fotointerpretação e da reambulação de campo foi obtido pela restituição dos dois modelos fotogramétricos que abrangem a área dos experimentos. A base cartográfica obtida pela restituição pode ser visualizada na Figura 26.

O mapa apresentado na Figura 26 é temático. O nível de medida das classes temáticas é qualitativo e a variável visual empregada é o tom de cor.

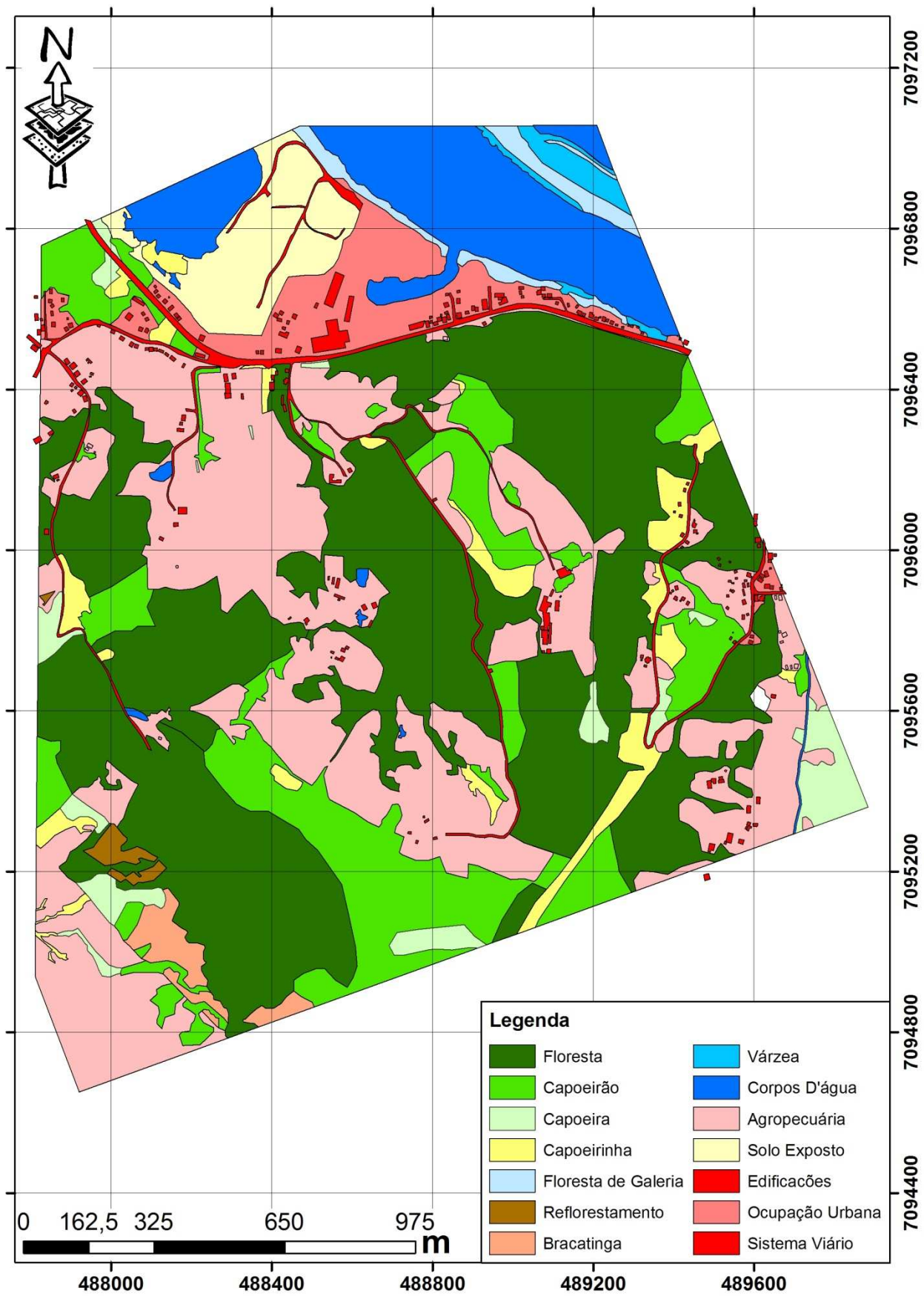


FIGURA 26 - BASE CARTOGRÁFICA OBTIDA PELA RESTITUIÇÃO

As classes temáticas obtidas pelo processo de restituição são:

- Edificações;
- Corpos d'água;
- Várzea;
- Floresta de galeria;
- Floresta;
- Capoeirão;
- Capoeira;
- Agropecuária (campo limpo e agricultura anual);
- Capoeirinha (agricultura abandonada).
- Bracatinga;
- Reflorestamento de *Pinnus sp.*;
- Solo exposto.

Ressalta-se mais uma vez que as edificações, Corpos d'água, solo exposto e agropecuária não são classes da cobertura vegetal, mas fazem parte da área dos experimentos. Estas classes foram incluídas na classificação com o intuito de verificar o desempenho da metodologia proposta para classes distintas das formações vegetais.

As áreas ocupadas por cada classe temática foram totalizadas e são apresentadas na Tabela 2.

TABELA 2 - TOTALIZAÇÃO DAS CLASSES TEMÁTICAS

CLASSE TEMÁTICA	ÁREA (ha)	Percentual(%)
1. Edificações	3,552	1,03
2. Corpos d'água	26,928	7,81
3. Várzea	2,752	0,80
4. Agropecuária	96,800	28,09
5. Bracatinga	3,318	0,96
6. Capoeirinha	10,795	3,13
7. Floresta de galeria	4,032	1,17
8. Floresta	130,696	37,92
9. Capoeirão	43,095	12,50
10. Capoeira	9,124	2,65
11. Reflorestamento	1,235	0,36
12. Solo exposto	12,117	3,57
<b>Total</b>	<b>344,646</b>	<b>100,00</b>

A maior parte da área de estudo, ou seja 57,32%, é ocupada por formações naturais: floresta, capoeirão, capoeira, floresta de galeria e várzea, além de

reflorestamento e bracatinga. As áreas ocupadas por agropecuária, capoeirinha e solo exposto respondem por 34,79% da área de estudo. Com relação aos rios e lagos (Corpos d'água), o percentual ocupado por estas classes foi de 7,81%. As edificações respondem por apenas 1,03% da área de estudo.

Se a classificação fosse realizada somente em relação à cobertura vegetal não seriam verificados 43,68% da área pesquisada. Esta foi mais uma motivação para a inserção das classes não pertencentes à cobertura vegetal.

#### 4.2.2 Determinação dos Perfis Longitudinais

Os resultados referentes à altimetria foram 218 perfis longitudinais, espaçados de 10 metros. Estes perfis são constituídos por 604.852 pontos.

A partir dos pontos cotados foi interpolada no programa Surfer 8.0 a grade regular para a área de estudo. Existem vários métodos de interpolação que podem ser utilizados para auxiliar no refinamento de dados amostrais, possibilitando a geração do MDT. O vizinho mais próximo é o método mais simples e tem como principal característica assegurar que o valor interpolado seja um dos valores originais, ou seja, não gera novos valores (BOTELHO *et al.*, 2005). De acordo com estes autores o vizinho mais próximo gerou um MDT mais próximo da superfície real do terreno.

A Figura 27 apresenta os perfis longitudinais obtidos pela restituição fotogramétrica.



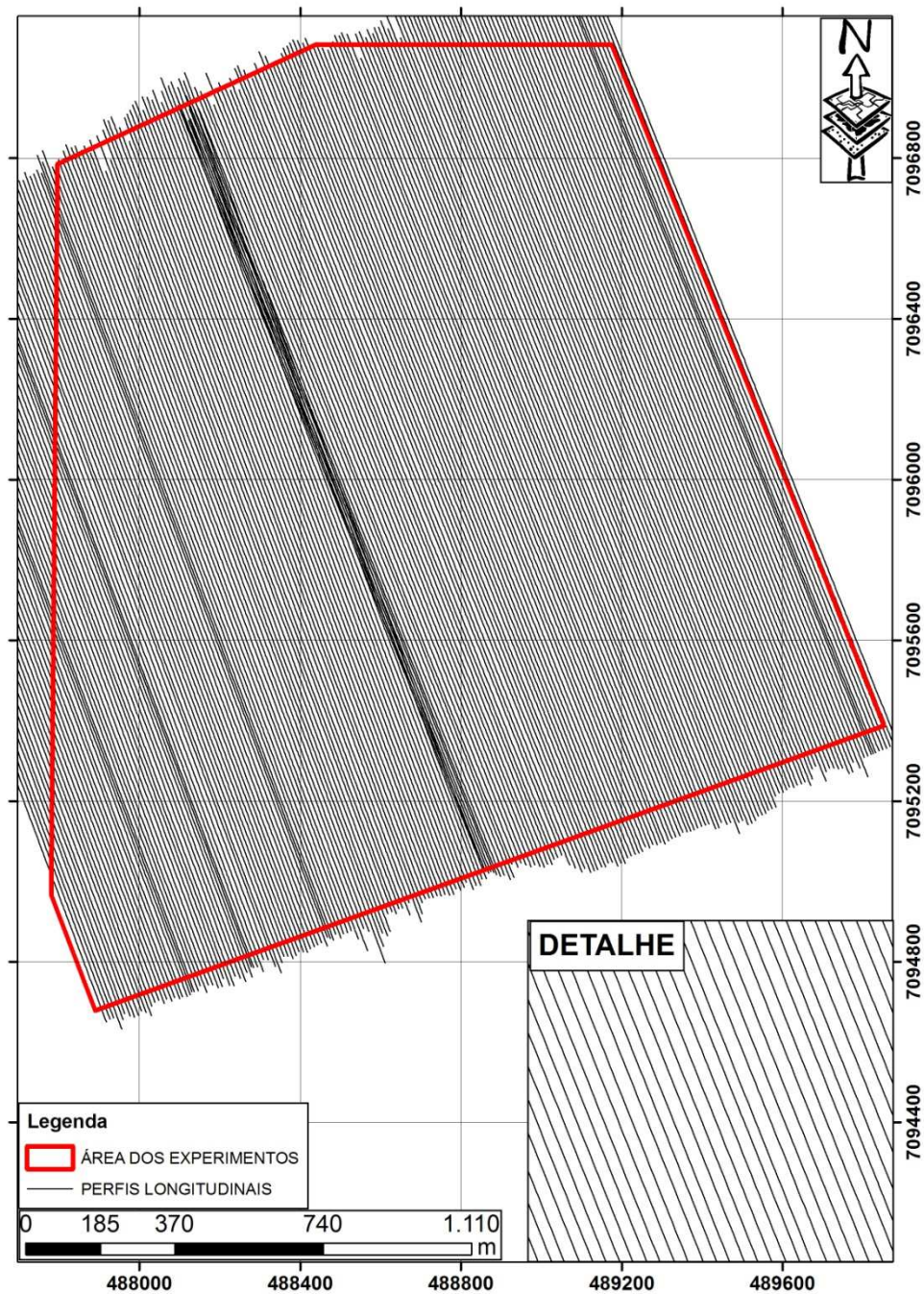


FIGURA 27 - PERFIS LONGITUDINAIS PROVENIENTES DA RESTITUIÇÃO FOTOGRAMÉTRICA

A análise da Figura 27 permite afirmar que a amostragem altimétrica abrange toda a área dos experimentos, não sendo perceptível a existência de vazios, isto é, locais onde não foi procedida a amostragem da superfície topográfica pelo método fotogramétrico. O detalhe apresentado na figura 26 detalha as linhas de varredura, possibilitando a visualização do espaçamento entre as respectivas linhas.



#### 4.2.3 Compatibilidade entre as coordenadas do mapa temático e a imagem de intensidade

O mapa temático obtido pela restituição planialtimétrica e reambulação de campo foi utilizado como referência para a realização de comparações e determinação de índices de acertos, tanto do filtro morfológico (informações altimétricas) quanto para a classificação da cobertura do solo empregando as funções discriminantes.

Apesar de estarem referidos ao mesmo sistema de coordenadas WGS 84, poderiam ocorrer diferenças entre as coordenadas determinadas pela varredura LASER e as coordenadas determinadas pela fotogrametria. Portanto, foi analisada a compatibilidade geométrica dos dados provenientes da fotogrametria e os dados LASER, mediante a comparação das coordenadas de pontos identificáveis nas duas fontes de dados.

Para a comparação foram utilizadas a base cartográfica obtida pela restituição fotogramétrica e a imagem de intensidade derivada a partir dos dados LASER.

A imagem de intensidade foi obtida através da interpolação dos pontos LASER, utilizando-se o programa Terrascan. A resolução geométrica da imagem de intensidade é de 1 metro.

A visualização e pesquisa das coordenadas foi obtida através do programa Arcgis 9.2. Os vetores obtidos pela fotogrametria foram sobrepostos à imagem de intensidade e então foram selecionados 10 pontos distribuídos por toda a área dos experimentos. Foram escolhidos pontos bem definidos, citando como exemplos, intersecção de vias e bordas de áreas agrícolas. As coordenadas obtidas nesta atividade e as discrepâncias são apresentadas na Tabela 3.

TABELA 3 - COMPARATIVO DE COORDENADAS LASER x FOTOGRAMETRIA

Ponto	Fotogrametria		LASER		ERROS		
	E <sub>f</sub> (m)	N <sub>f</sub> (m)	E <sub>l</sub> (m)	N <sub>l</sub> (m)	ΔE(m)	ΔN(m)	Δ(m)
1	488583,561	7096503,877	488583,927	7096503,314	-0,366	0,563	0,672
2	488043,813	7096592,325	488044,250	7096591,81	-0,437	0,515	0,675
3	489239,899	7096564,789	489240,117	7096564,299	-0,218	0,490	0,536
4	489117,768	7095928,844	489117,500	7095929,412	0,268	-0,568	0,628
5	488639,303	7095937,369	488639,020	7095937,9	0,283	-0,531	0,602
6	487856,792	7095878,151	487856,532	7095878,475	0,260	-0,324	0,415
7	489679,001	7095796,152	489679,315	7095795,885	-0,314	0,267	0,412
8	489141,399	7095448,54	489141,956	7095448,83	-0,557	-0,290	0,628
9	488746,500	7095292,487	488746,650	7095292,47	-0,150	0,017	0,151
10	489072,736	7096963,972	489072,485	7096963,183	0,251	0,789	0,828
$\bar{x}$					-0,098	0,093	0,555
$\sigma$					0,332	0,397	0,188
EMQ					0,330	0,411	0,483

Na Tabela 3:

$$\Delta E = E_f - E_l \quad (119)$$

$$\Delta N = N_f - N_l \quad (120)$$

$$\Delta = \sqrt{\Delta E^2 + \Delta N^2} \quad (121)$$

O erro médio quadrático (EMQ) permite avaliar a precisão de uma série de observações, pois apresenta a vantagem de que erros maiores apresentam maior influência devido os seus quadrados.

O valor de EMQ é determinado através do uso da equação (122):

$$EMQ = \sqrt{\frac{\sum \Delta^2}{n}} \quad (122)$$

Na equação,  $n$  é o número de observações, ou seja, 10.

Ao analisar a Tabela 3 verifica-se que a discrepância planimétrica média foi de 55,5 centímetros e o desvio padrão de 18,8 centímetros. O EMQ resultou em 48,3 centímetros. E ainda conhecida a precisão nominal fornecida pelo fabricante do *laser scanner*, que é de 50 centímetros.

Confrontando o erro médio quadrático obtido para os 10 pontos analisados com a precisão nominal do sistema, pode-se afirmar que existe compatibilidade geométrica entre as medidas derivadas pelo processo fotogramétrico e *laser scanner*, tornando possível o estabelecimento da comparação entre as duas fontes de dados.

#### 4.3 DEFINIÇÃO DO MODELO DIGITAL DO TERRENO

São apresentados nesta seção os resultados obtidos na definição da superfície topográfica para a área dos experimentos. Para mostrar o desempenho do processo de filtragem dos pontos foram escolhidas duas áreas menores, ocupadas com diferentes tipos de cobertura vegetal. A análise para toda a área de estudo também é apresentada.

##### 4.3.1 Resultados da Filtragem para a Área 1

A primeira área escolhida corresponde a uma transição entre formação natural média e agropecuária com a presença de árvores isoladas, edificações e lagos. As coordenadas limitantes (em metros) desta área são 488.080; 7.095.800 e 488.440; 7.096.400. A Figura 28 apresenta o anaglifo para a área 1.

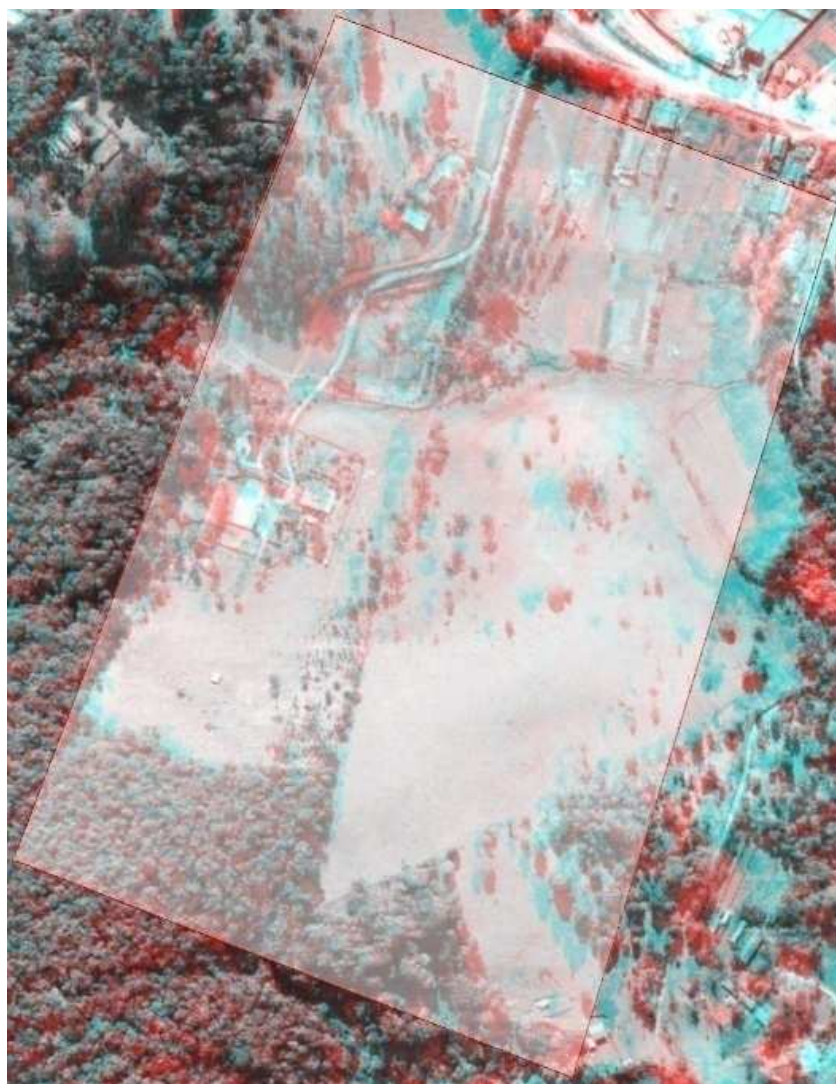


FIGURA 28 - ANAGLIFO DA ÁREA 1

O número de pontos LASER para esta área foi de 393.344 pontos (primeiro e último retorno). A área de estudo totalizou 14,40 hectares, o que resultou em uma densidade média de 2,72 pontos por metro quadrado.

Empregando-se a metodologia apresentada na seção 3.3.5, ao realizar a primeira iteração do filtro morfológico adaptado, do total de 393.344 foram selecionados 110.469 pontos como pertencentes ao terreno. Isto equivale a selecionar 28,08 % dos pontos como sendo pertencentes à superfície topográfica.

Na segunda iteração dos 393.344, 92.683 pontos foram selecionados como pertencentes ao terreno, ou seja, 23,56 % dos pontos.

Na terceira iteração foram selecionados 76.401, o que equivale a 19,42% dos pontos originais. Nesta iteração o critério de parada foi alcançado para toda a área de estudo.

A Figura 29 apresenta as grades regulares obtidas em cada iteração, com a filtragem dos pontos para a área 1.

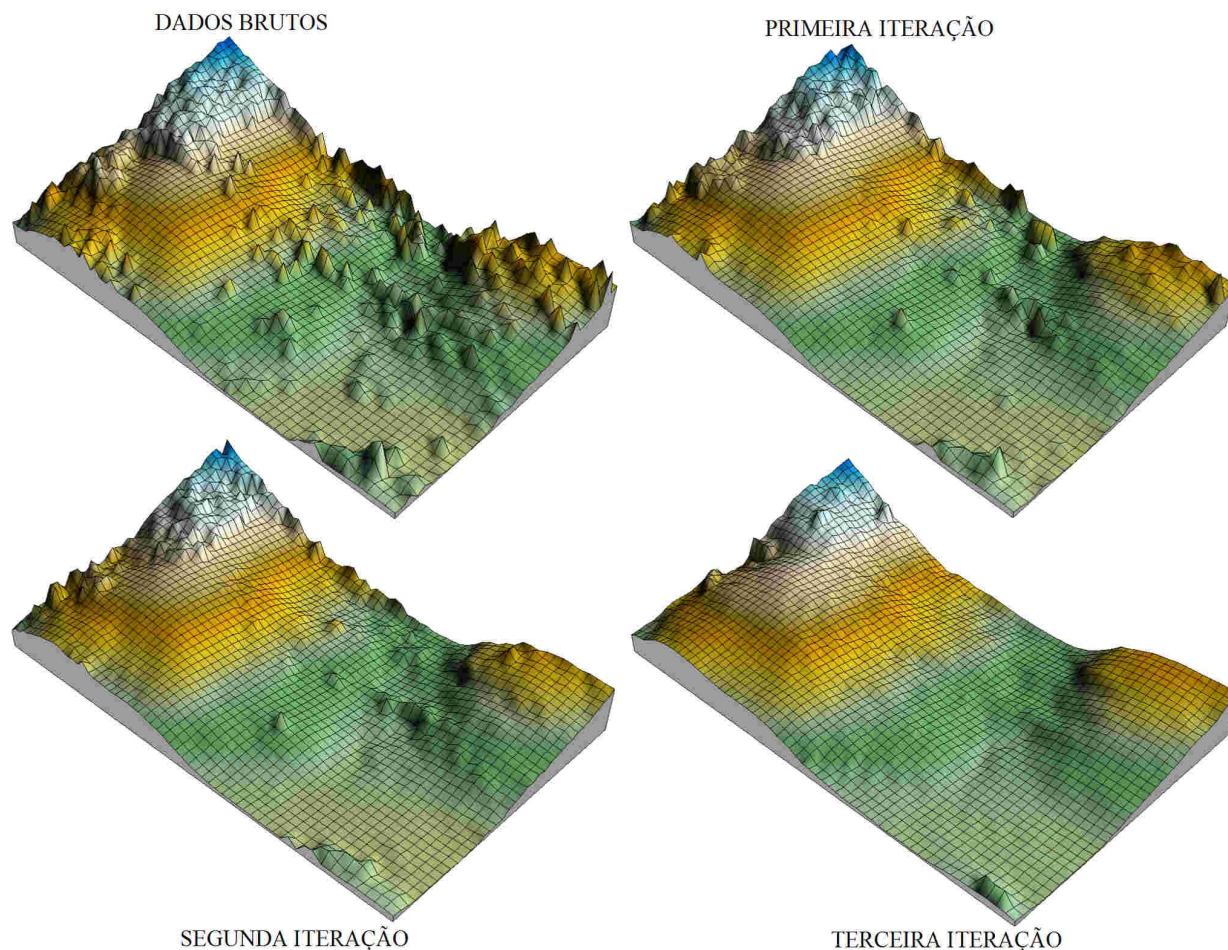


FIGURA 29 - ITERAÇÕES DO FILTRO MORFOLÓGICO PARA A ÁREA 1

A análise visual das grades regulares obtidas para a área 1 mostra que o resultado do filtro morfológico é bastante satisfatório. Percebem-se, no entanto, a presença de pontos acima do terreno após a conclusão do processo de filtragem e que podem ser observados na 3ª iteração.

A densidade média de pontos separados como pertencentes à superfície topográfica foi de 0,6 ponto/m<sup>2</sup> para a área de teste 1.

A grade regular obtida na terceira iteração foi então comparada visualmente com a grade regular gerada a partir das observações fotogramétricas. Esta análise é mostrada na Figura 30.



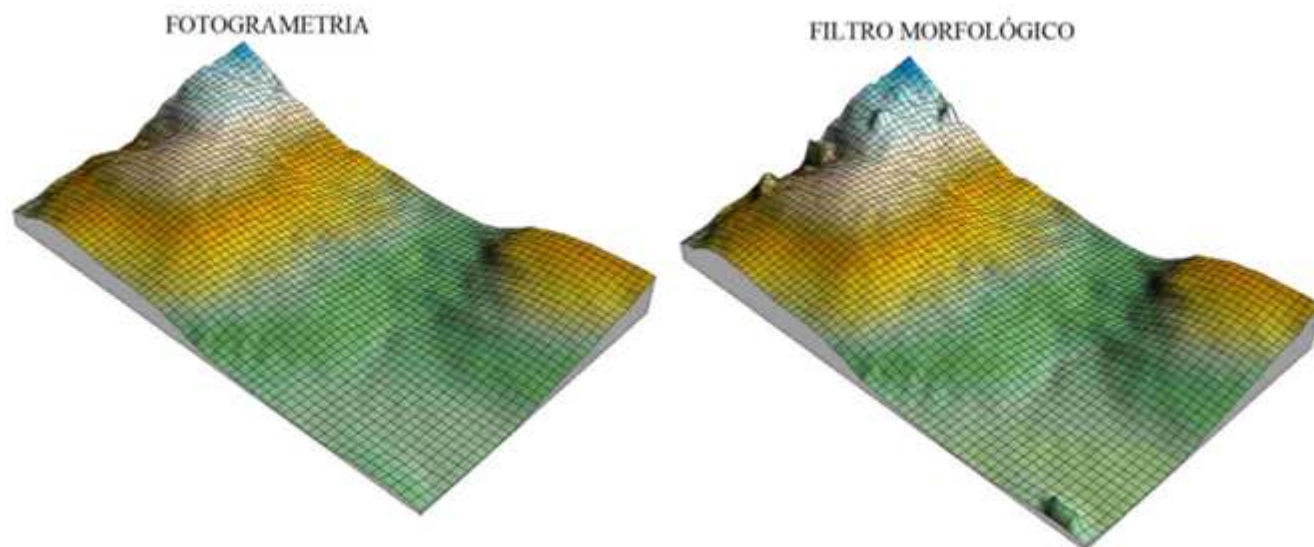


FIGURA 30 - FOTOGRAMETRIA x FILTRO MORFOLÓGICO NA ÁREA 1

A análise visual da Figura 30 possibilita afirmar que o filtro morfológico apresentou um resultado bastante satisfatório quando comparado com a fotogrametria.

Adicionalmente, foram geradas as curvas de nível empregando-se os pontos filtrados e os provenientes da fotogrametria. A separação vertical das curvas de nível escolhida foi de 5 metros. A Figura 31 possibilita a visualização destas curvas.

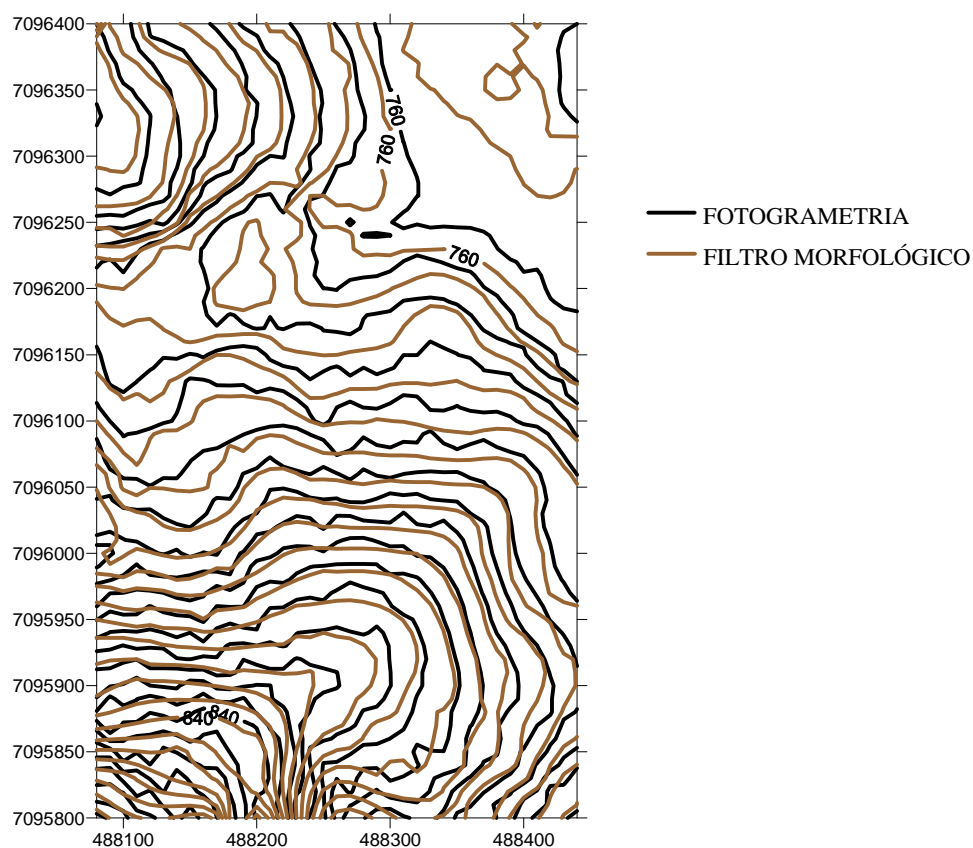


FIGURA 31 - CURVAS DE NÍVEL PARA A ÁREA 1

A análise visual da Figura 31 mostra a boa conformidade entre as duas fontes de dados, não sendo perceptíveis grandes discrepâncias na comparação.

Outro teste realizado obteve a matriz de diferenças altimétricas. Esta matriz armazenou a diferença entre a grade regular gerada a partir dos pontos filtrados e a grade obtida através dos pontos fotogramétricos.

A obtenção da matriz de diferenças forneceu os seguintes valores estatísticos:

- Diferença média entre as altitudes: -1,294 metro;
- Desvio padrão das diferenças: 1,603 metro;
- Maior diferença de altitude: 6,503 metros;
- Menor diferença de altitude: -6,851 metros;
- EMQ: 2,125 metros.

A análise dos resultados estatísticos permite afirmar que, quando comparada com a superfície gerada a partir das observações derivadas pelo processo fotogramétrico, a superfície obtida a partir dos pontos separados pelo filtro morfológico está rebaixada de um valor médio de 1,294 metro. Considerando-se o EMQ, a discrepância entre as duas superfícies obtidas é de 2,125 metros.

#### 4.3.2 Resultados da Filtragem para a Área 2

A área 2 foi selecionada em um local mais íngreme, sendo delimitada pelas coordenadas (em metros) 489.000; 7.096.000 e 489.250; 7.096.500. A Figura 32 apresenta o anaglifo referente à área 2.

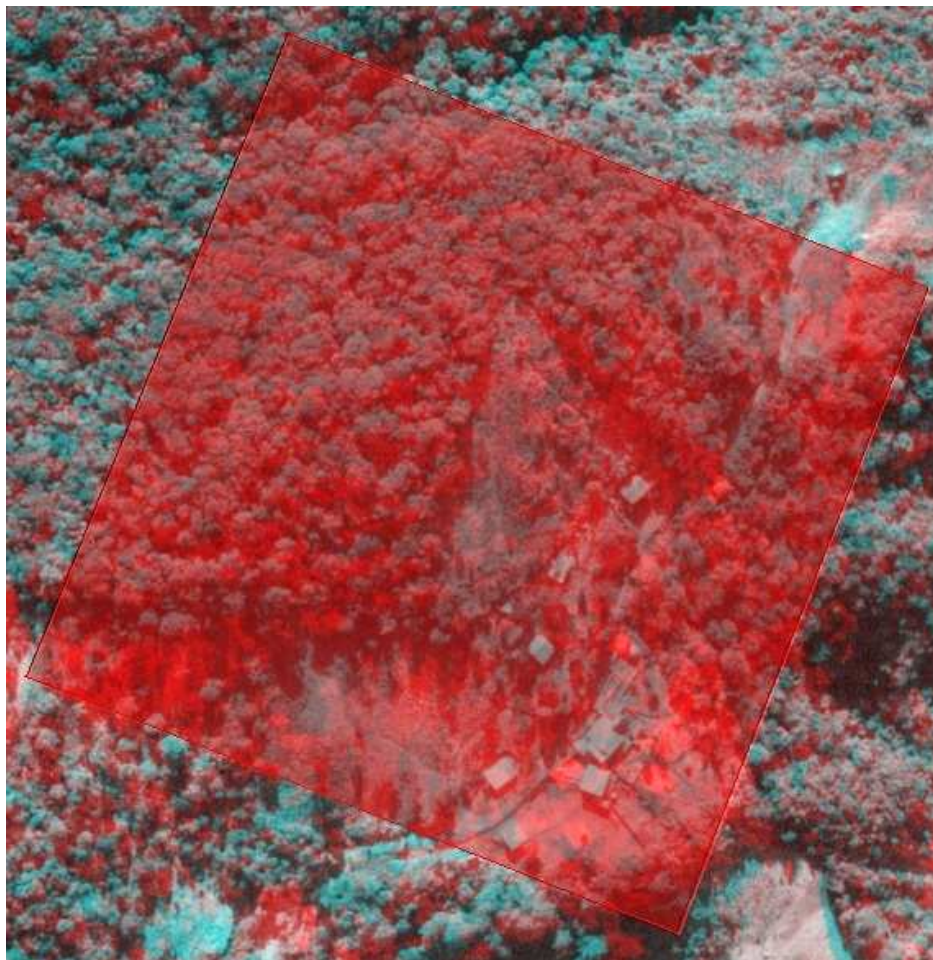


FIGURA 32 - ANAGLIFO DA ÁREA 2

A área 2 apresenta formações naturais médias e avançadas, além de áreas de capoeirinha e edificações isoladas.

O número de pontos LASER na área 2 foi de 131.149, distribuídos em 6,25 ha. A densidade média de pontos para a área 2 é de 2,09 por metro quadrado.

Com relação ao filtro morfológico para esta área foram obtidos os seguintes resultados:

- Primeira iteração: do total de 131.149, foram filtrados 23.105 pontos como pertencentes à superfície topográfica;
- Segunda iteração: 21.842 pontos filtrados;
- Terceira iteração: 20.802 pontos selecionados como pertencentes à superfície topográfica. Ou seja, 15,86% dos pontos foram selecionados para a área 2.

A densidade média de pontos na superfície topográfica é de 0,33 por metro quadrado.



As iterações do filtro morfológico para a área 2 podem ser visualizadas na figura 33.

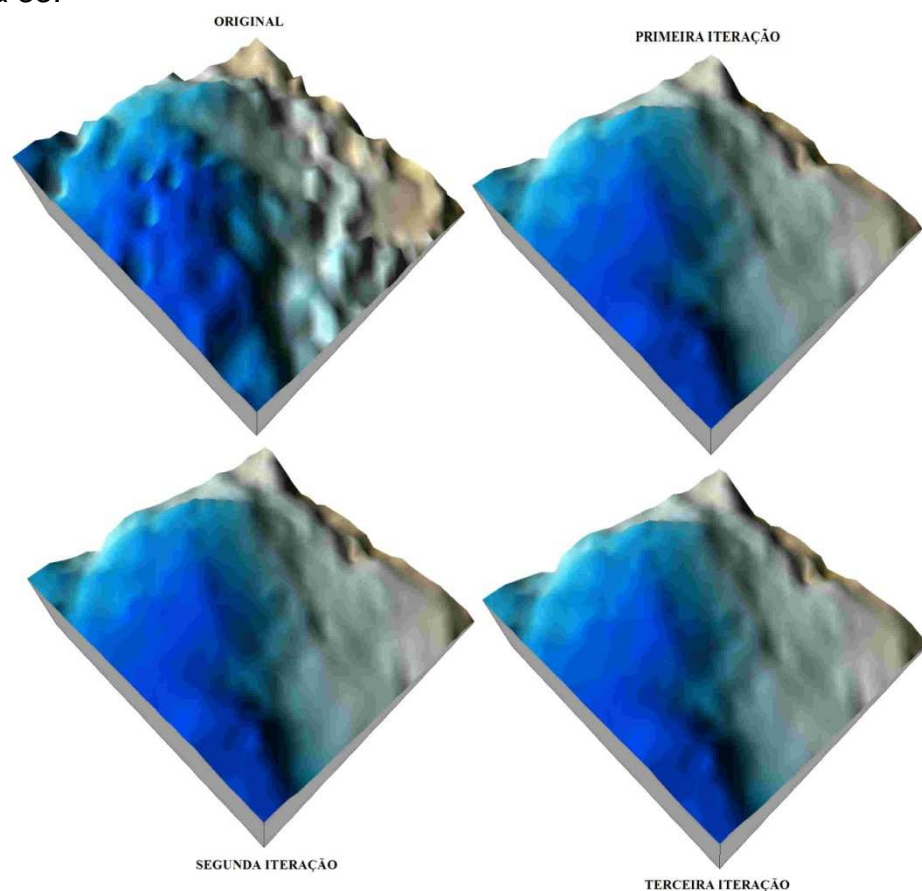


FIGURA 33 - ITERAÇÕES DO FILTRO MORFOLÓGICO PARA A ÁREA 2

A partir dos pontos filtrados na terceira iteração foi interpolada a grade regular e comparada visualmente com a grade obtida a partir das observações fotogramétricas. Esta comparação visual é apresentada na Figura 34.

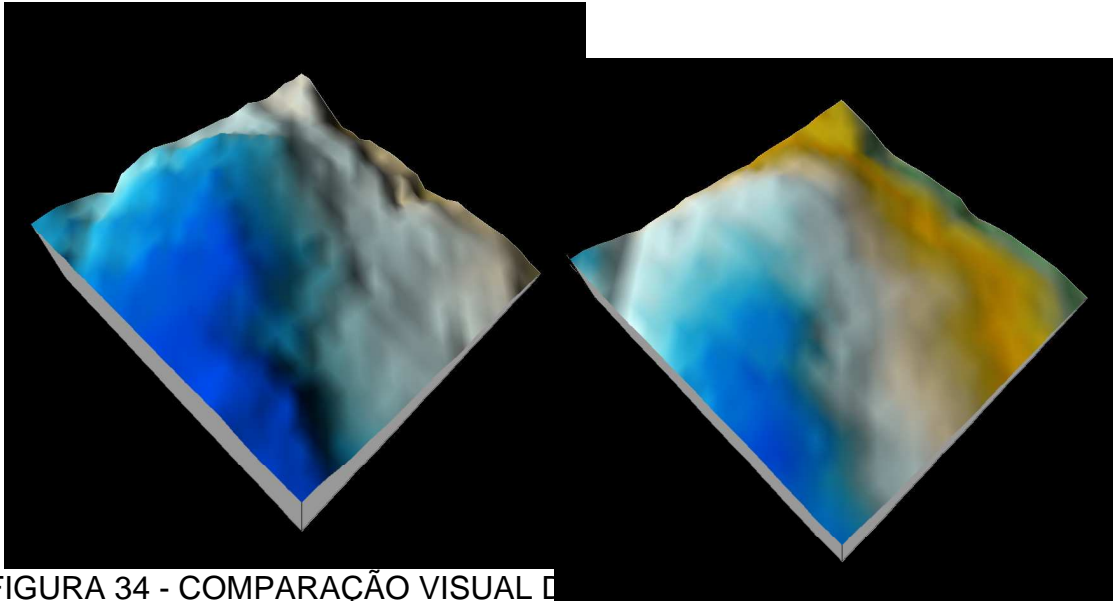


FIGURA 34 - COMPARAÇÃO VISUAL [

2

A análise visual da Figura 34 possibilita afirmar que a filtragem dos pontos LASER na área 2 apresentou um desempenho inferior quando comparada com a área 1. Isto pode ser visualizado nas discontinuidades da superfície obtida a partir dos pontos filtrados.

Foram obtidas as curvas de nível empregando-se os pontos filtrados e os provenientes da fotogrametria. A separação vertical das curvas de nível foi de 5 metros. A Figura 35 possibilita a visualização das curvas de nível geradas a partir destas duas fontes de dados.

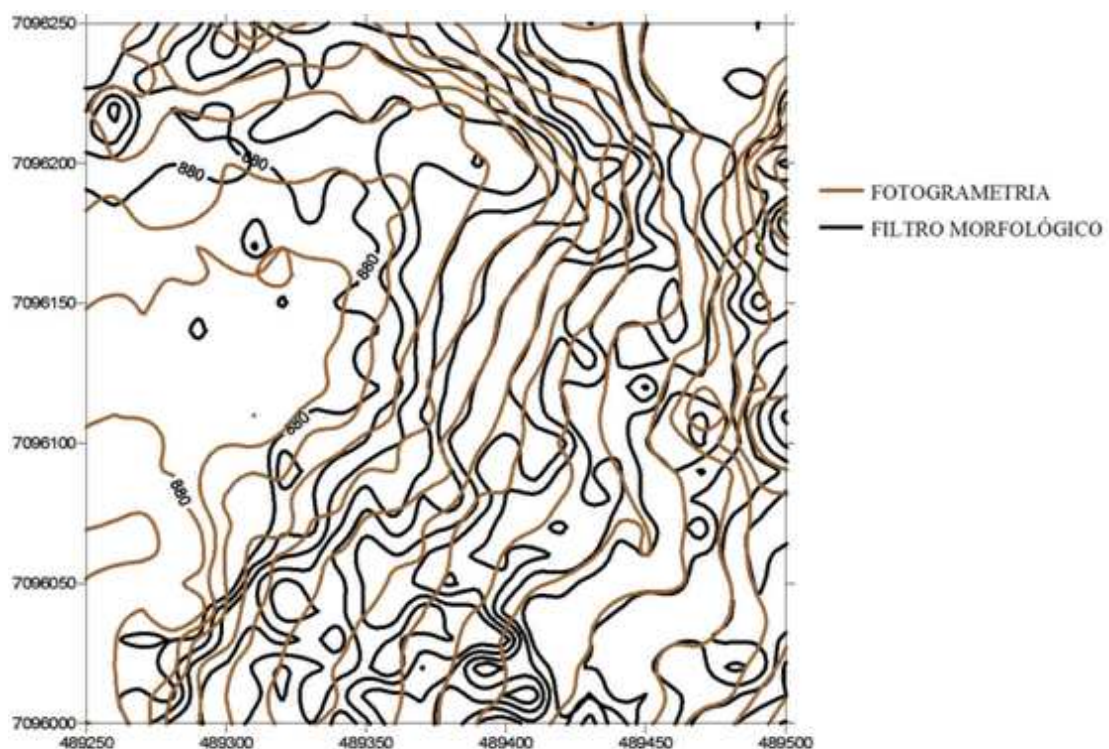


FIGURA 35 - CURVAS DE NÍVEL PARA A ÁREA 2

Pode-se concluir pela análise da Figura 35 que o filtro morfológico não obteve sucesso na separação dos pontos da área 2. É possível perceber que ainda restaram vários pontos acima da superfície topográfica, sendo esta situação visualizada através da formação de curvas de nível fechadas, distribuídas por toda a área analisada.

De forma análoga ao procedido com a área 1, obteve-se a matriz de diferenças altimétricas entre a grade gerada a partir dos pontos LASER e a grade gerada a partir das observações fotogramétricas. Foram obtidos os seguintes valores estatísticos para esta matriz:

- Diferença média entre as altitudes: -2,653 metros;
- Desvio padrão das diferenças: 2,287 metros;
- Maior diferença de altitude: 5,985 metros;
- Menor diferença de altitude: -8,652 metros;
- EMQ: 3,503 metros.

Os resultados estatísticos apresentados para as diferenças altimétricas na área 2 refletem o pior desempenho neste segundo experimento. O rebaixamento da superfície obtida com os pontos filtrados em relação à superfície gerada a partir das observações fotogramétricas aumentou para 2,653 metros. O valor do EMQ reflete o pior desempenho do filtro morfológico na área 2 com discrepância de 3,503 metros.

#### 4.3.3 Resultados da Filtragem para a Área Total

Nesta seção os experimentos foram processados para toda a área de estudo. O número total de pontos, correspondente ao primeiro e último retorno, é de 6.452.347 pontos. A densidade média é de 1,87 ponto por metro quadrado.

Os resultados para as iterações são apresentados a seguir:

- Primeira iteração: filtrados 1.777.723 pontos no solo;
- Segunda iteração: 1.681.077 pontos no solo;
- Terceira iteração: 1.661.503 pontos no solo.

Foram filtrados como pontos no solo 25,72% do total de pontos. Isto equivale a uma densidade média de 0,48 ponto por metro quadrado.

Na Figura 36 apresentam-se as grades regulares obtidas a partir dos dados fotogramétricos e dos dados LASER filtrados.

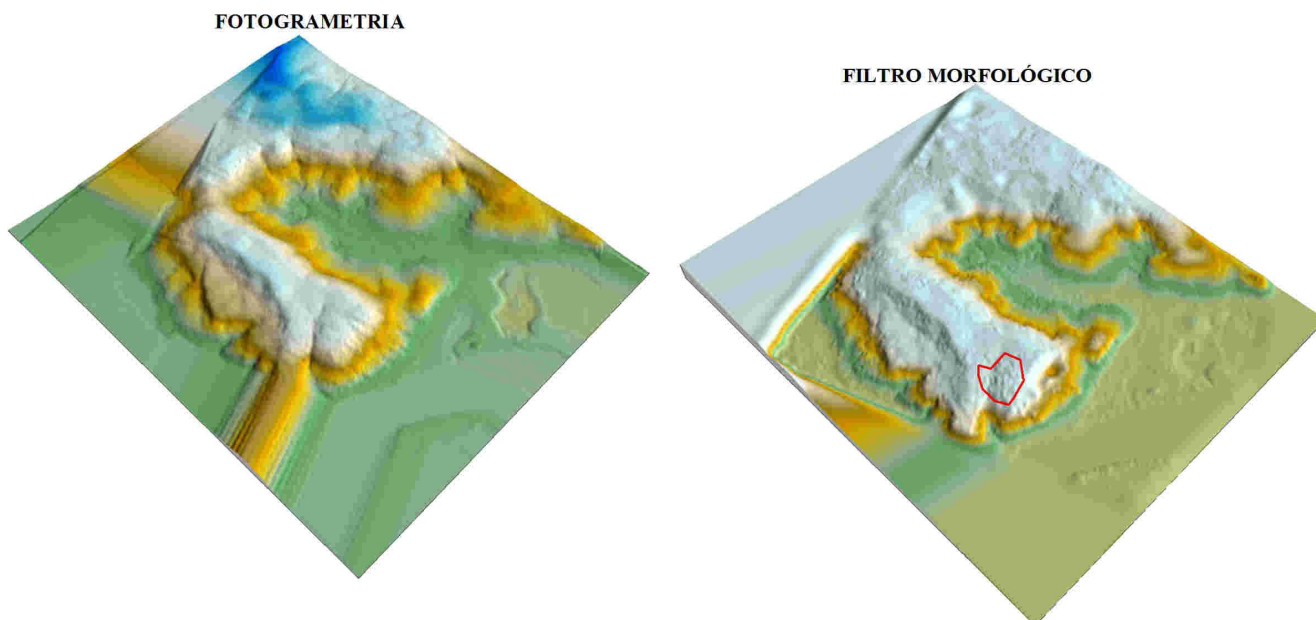


FIGURA 36 - GRADES: FOTOGRAMETRIA x FILTRO MORFOLÓGICO

A análise visual da grade regular gerada a partir dos pontos filtrados possibilita detectar ainda locais em que ocorrem descontinuidades na superfície topográfica. Uma destas descontinuidades é destacada na Figura 36.

O filtro morfológico adaptado foi desenvolvido para aplicações em áreas íngremes cobertas por vegetação e no local escolhido foram detectados outros tipos de cobertura vegetal e por isso o processo proposto não conseguiu filtrar adequadamente os pontos pertencentes à superfície topográfica. Zaksek e Pfeifer (2006) apontam a necessidade da utilização de um segundo tipo de filtro para ser empregado em conjunto com o filtro morfológico, isto é, de forma complementar.

A matriz de diferenças gerada a partir das grades LASER filtrada e da grade fotogramétrica apresentou os seguintes valores estatísticos:

- Número de elementos: 52.961;
- Diferença média: -1,518 m;
- Desvio padrão: 2,901 m;
- Menor diferença: -9,136 m;
- Maior diferença: 5,978 m;
- EMQ: 3,275 m.

Utilizando um desvio padrão de  $1\sigma$  determina-se um intervalo para análise das diferenças que inicia em -4,933 e vai até 1,497 metros. Neste intervalo foram detectadas 82,24% das diferenças calculadas, ou seja, 38.259 células com diferenças contidas no intervalo de erro padrão.

Para um intervalo de -1 a 1 metro foram detectadas 52,24% das diferenças, o que equivale a 27.667 células.

Considerando-se ainda os valores estatísticos obtidos percebe-se a tendência do rebaixamento da superfície gerada a partir dos pontos separados pelo filtro morfológico quando comparada com a superfície obtida a partir das observações fotogramétricas.

A partir da matriz de diferenças de altura (em metros) obteve-se a carta temática mostrada na Figura 37.

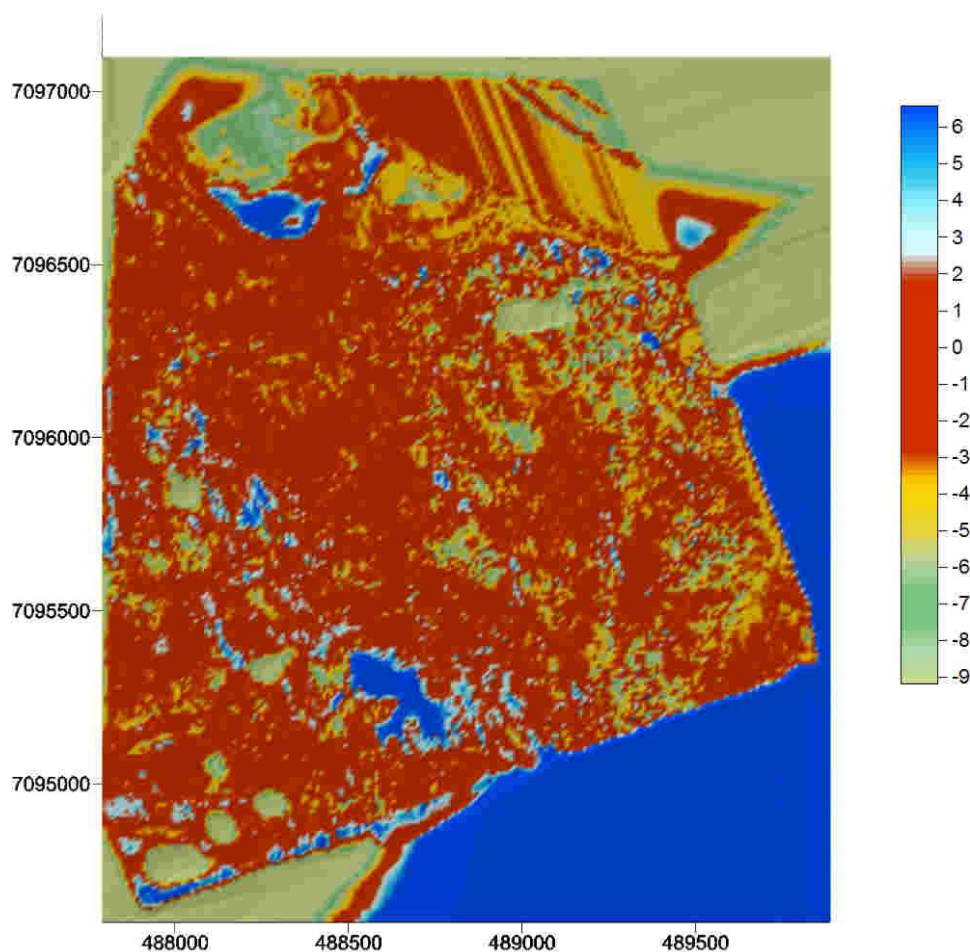


FIGURA 37 - MATRIZ DE DIFERENÇAS DE ALTURA PARA A ÁREA TOTAL

#### 4.4 DETERMINAÇÃO DA COBERTURA VEGETAL ATRAVÉS DA CLASSIFICAÇÃO DAS GRADES REGULARES

Para se avaliar a metodologia apresentada, foram necessários mais de duas dezenas de experimentos com as grades regulares até se obter os resultados finais expostos. As três grades regulares foram selecionadas visando atestar o desempenho dos classificadores para resoluções geométricas distintas. Este esforço ocorreu no sentido de verificar o funcionamento da metodologia proposta, tendo em mente que os sistemas LIDAR atuais tem a capacidade de obter nuvens de pontos com densidades de até 20 pontos/m<sup>2</sup> e vislumbrando também a sua utilização também na classificação de imagens digitais de alta resolução geométrica.

Conforme apresentando na metodologia foram criadas três situações para a classificação da nuvem de pontos:

- Grade regular com resolução de 10 metros;
- Grade regular com resolução de 5 metros;
- Grade regular com resolução de 1 metro.

Cada uma das grades foi classificada empregando-se os classificadores Fisher (linear), Quadrático e Regressão Logística Oculta (RLO).

A nomenclatura das formações naturais e antrópicas adotada para o experimentos que seguem são:

- AP - Agropecuária;
- CS - Capoeirinha ;
- SO - Solo exposto;
- ED- Edificações;
- VA - Várzea;
- BR - Bracatinga;
- RE - Reflorestamento;
- AV - Floresta;
- ME - Capoeirão;
- IN - Capoeira;
- CI - Floresta de galeria;
- AG - Corpos d'água.



Com relação ao processo de amostragem conduzido para as grades regulares destaca-se que foram obtidos a seguinte quantidade  $n$  de amostras:

- grade regular de 10 metros de resolução : 218 amostras;
- grade regular de 5 metros de resolução : 840 amostras;
- grade regular de 1 metro de resolução : 34.123 amostras.

A partir das amostras foram determinados as funções discriminantes. Os coeficientes destas funções são apresentados visando tornar possível a verificação dos resultados alcançados.

Os resultados da classificação foram armazenados no formato matricial, isto é, em cada célula das matrizes obtidas com os experimentos de classificação está a codificação numérica das tipologias referentes a cobertura vegetal para a célula classificada.

A base cartográfica obtida pelos processos fotogramétricos também foi armazenada no formato matricial.

A matriz de confusão para a classificação foi obtida através da totalização das classes armazenadas na matriz obtida com a classificação e a matriz gerada a partir da base cartográfica. Ela apresenta como dimensões 12 linhas por 12 colunas. Para facilitar o entendimento sobre como a matriz de confusão é obtida considere a linha 125 e a coluna 140. Ao consultar a matriz classificada para a posição fornecida obtêm-se o valor 10. Nesta mesma posição na matriz obtida a partir da base cartográfica obtêm-se o valor 11. Na linha 10 e coluna 11 da matriz de confusão será adicionado um. No final do processo de criação da matriz de confusão são totalizados as células que compõe as colunas da matriz. Esta totalização será empregada para dividir as células contidas em uma mesma coluna e com isso obter um valor percentual relativo ao acerto ou erro da classificação. A disposição da matriz de confusão é mostrada abaixo:

$$\begin{bmatrix} \frac{T_{11}}{\sum T_1} & \frac{T_{12}}{\sum T_2} & \dots & \frac{T_{112}}{\sum T_{12}} \\ \frac{T_{21}}{\sum T_1} & \frac{T_{22}}{\sum T_2} & & \frac{T_{212}}{\sum T_{12}} \\ \vdots & \vdots & & \vdots \\ \frac{T_{121}}{\sum T_1} & \frac{T_{122}}{\sum T_2} & & \frac{T_{1212}}{\sum T_{12}} \end{bmatrix}$$

O termo  $\sum T_j$  refere-se ao somatório dos totais  $T_{ij}$  obtidos em cada coluna  $j$ .

A diagonal principal da matriz de confusão fornece o acerto relativo para a cada uma das doze classes de cobertura vegetal.

#### 4.4.1 Classificação Empregando a Função Discriminante de Fisher

Antes de proceder a classificação apresentada a seguir foram realizados testes para verificar o desempenho na classificação das células da grade regular de 10 metros empregando diferentes combinações de cofatores. A partir destes testes e observando o proposto na seção 3.3.8, foram selecionadas as três combinações de cofatores para as quais se procedeu a classificação:

- Primeiro teste: h e l;
- Segundo teste: h, l e altura da cobertura vegetal.
- Terceiro teste: foram empregadas todas as variáveis para a classificação.

As combinações foram testadas com o algoritmo *stepwise forward*, cujo teste mostrou que as variáveis escolhidas tem relevância na classificação com valores estatístico de P menores que 0,05, ou seja, são estatisticamente significantes ao nível de confiança de 95%.



#### 4.4.1.1 Grade regular com resolução de 10 metros empregando-se as variáveis h e I

O primeiro experimento apresentado mostra os resultados obtidos ao empregar-se a Função Discriminante Linear (FDL) e a combinação das variáveis altitude e intensidade média.

O ajuste, a partir das 218 amostras, forneceu duas equações de discriminação linear, cujas variáveis de entrada são a Altitude Média (h) e a Intensidade Média (I) na célula. As equações de discriminação deste primeiro experimento são mostradas abaixo:

$$\hat{Y}_1 = -74,41030 + 0,09061.h - 0,00405.I \quad (123)$$

$$\hat{Y}_2 = 1,10552 - 0,00739.h + 0,05356.I \quad (124)$$

As equações (123) e (124) possibilitaram a determinação dos centróides para cada uma das 12 classes temáticas. Os centróides com dimensão canônica igual a dois são apresentados na Tabela 4.

TABELA 4 - CENTRÓIDES DAS CLASSES TEMÁTICAS PARA A GRADE 10m - VARIÁVEIS I E H

Classe	$\hat{Y}_1$	$\hat{Y}_2$
AP	-3,22216	1,92437
AV	10,21990	-0,48802
BR	3,63155	0,87865
CS	-6,23229	-0,41630
CI	-5,68311	-2,53160
ED	-7,39482	-1,56486
IN	-6,87167	-0,76279
ME	7,00076	-0,02433
RE	4,09542	0,06687
AG	6,39557	0,08127
SO	9,23706	-1,07917
VA	-6,64929	0,58908

Concluído o ajuste foi possível implementar a classificação da grade de 10 metros empregando-se como variáveis a Altitude Média (h) e a Intensidade Média (I) na célula.

Após a classificação da grade regular de 10 metros obteve-se a matriz de confusão, conforme descrito na seção 3.3.9.4. O erro calculado e apresentado nesta tabela refere-se ao erro de omissão (análise da linha da matriz de confusão). Este resultado é mostrado na Tabela 5.

TABELA 5 - MATRIZ DE CONFUSÃO PARA A GRADE 10 m - VARIÁVEIS I E H

Classe	AP	BR	CS	CI	ED	AG	VA	AV	ME	IN	RE	SO	Soma
AP	0,679	0,000	0,105	0,000	0,022	0,000	0,011	0,000	0,000	0,000	0,018	0,166	1,000
BR	0,000	0,686	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,314	0,000	1,000
CS	0,084	0,000	0,636	0,000	0,009	0,007	0,000	0,000	0,000	0,246	0,002	0,016	1,000
CI	0,000	0,000	0,000	0,627	0,000	0,000	0,000	0,035	0,195	0,143	0,000	0,000	1,000
ED	0,352	0,000	0,000	0,000	0,193	0,008	0,022	0,038	0,014	0,029	0,000	0,344	1,000
AG	0,000	0,000	0,000	0,000	0,000	0,571	0,423	0,000	0,002	0,003	0,000	0,000	1,000
VA	0,000	0,000	0,000	0,000	0,000	0,398	0,602	0,000	0,000	0,000	0,000	0,000	1,000
AV	0,000	0,000	0,000	0,092	0,000	0,000	0,000	0,646	0,153	0,109	0,000	0,000	1,000
ME	0,000	0,000	0,000	0,048	0,000	0,001	0,004	0,199	0,589	0,159	0,000	0,000	1,000
IN	0,002	0,000	0,181	0,095	0,000	0,000	0,000	0,117	0,059	0,546	0,000	0,000	1,000
RE	0,000	0,049	0,000	0,000	0,000	0,000	0,000	0,074	0,000	0,000	0,877	0,000	1,000
SO	0,084	0,000	0,000	0,000	0,078	0,000	0,000	0,000	0,000	0,000	0,000	0,838	1,000

Para facilitar a visualização dos resultados mostrados na tabela 5 optou-se por obter o gráfico relativo aos acertos de cada classe e que podem ser conferidos na diagonal principal da matriz de confusão. Este gráfico é apresentado na Figura 38.

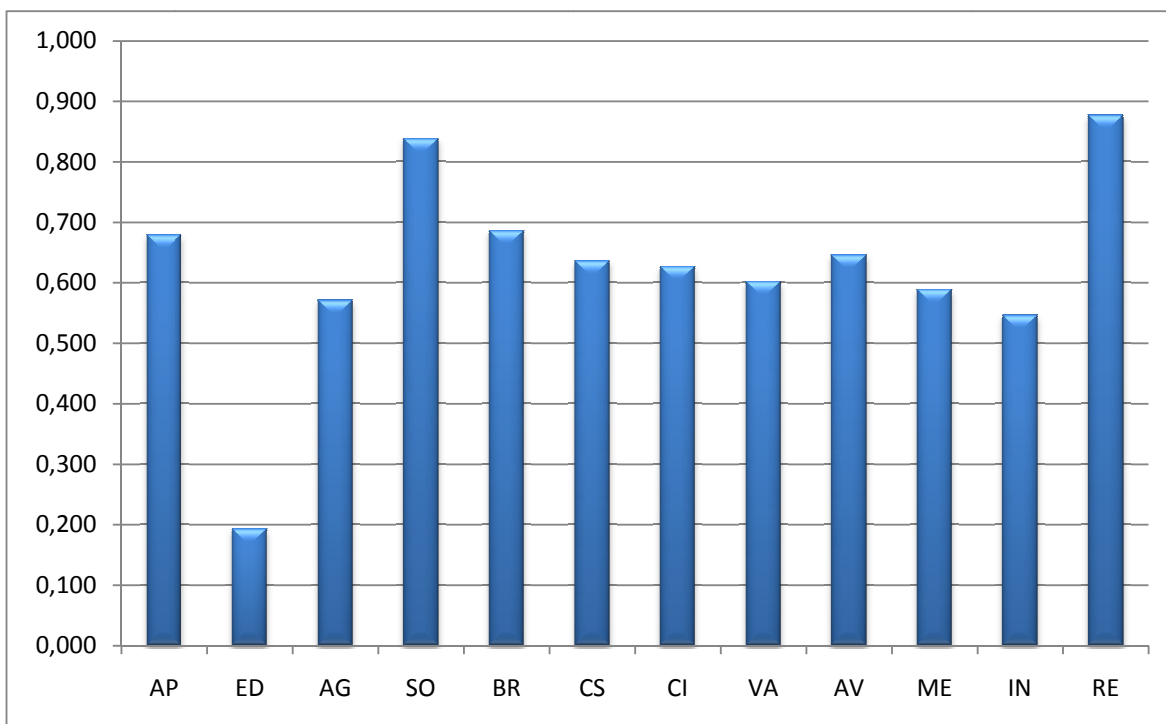


FIGURA 38 - ACERTO DA FDL COM COFATORES h E I

A partir da Tabela 5 e da Figura 38 as seguintes conclusões a respeito da classificação individual de cada classe foram obtidas:

- Agropecuária (AP): o acerto para a classe agropecuária foi de 67,4%. Ocorreu confusão com as classes capoeirinha (CS) de 10,5% devido ao fato de que esta classe é uma área de agropecuária abandonada e, portanto, apresenta similaridade com parâmetro de intensidade média. A maior confusão é verificada com a classe solo exposto (SO) com 16,6%. Neste caso, ao analisar a intensidade média desta classe verifica-se a proximidade de valores o que causou a confusão. Adiciona-se ainda o fato de que as áreas de solo exposto estão presentes em altitudes similares às das áreas ocupadas com agropecuária. Os índices confusão com a classes edificação (ED), reflorestamento (RE) e várzea (VA) foram pequenos sendo verificados os valores de 2,2%, 1,8% e 1,1% respectivamente. No caso da edificação (ED), a intensidade de resposta dos telhados de amianto confundiu o classificador, pois a intensidade média, neste caso, aproximou-se da resposta obtida com a agropecuária.

- Bracatinga (BR): o acerto da classificação da classe bracatinga foi de 68,6%. Ocorreu confusão apenas com a classe reflorestamento (RE). Nesta

situação, verifica-se que a altitude média das duas formações vegetais é bastante similar.

- Capoeirinha (CS): A classificação da formação vegetal capoeirinha obteve acerto de 63,6%. Ocorreu erro com a classe agropecuária (AP) de 8,4% que é próxima da confusão verificada para agropecuária (AP) versus capoeirinha (CS). A maior confusão pode ser percebida com a capoeira (IN) com 24,6%. Analisando a intensidade média desta classe percebe-se a semelhança entre as duas classes, além do que o capoeirinha (CS) é uma classe de sucessão vegetal que antecede a capoeira (IN) e esta proximidade acabou confundido o classificador. Ocorreram índices de confusão menor que 1% com as classes edificação (ED), corpos d'água (AG) e reflorestamento (RE). A confusão com o solo exposto (SO) foi de 1,6%. Analisando a imagem de intensidade percebeu-se a existência de regiões com intensidade semelhante nas duas classes.

- Floresta de galeria (CI): a floresta de galeria foi classificada corretamente com um índice de 62,7%. A maior confusão nesta classe ocorre com o capoeirão (ME) sendo verificado um índice de 19,5%, seguida da capoeira (14,3%) e floresta (3,5%). A análise da imagem de intensidade mostra valores de intensidade média próximos para as quatro classes o que causou o erro na separação desta três formações naturais.

- Floresta (AV): para esta classe o índice de acerto de 64,6%. Observa-se a confusão com a classes floresta de galeria (9,2%), capoeirão (15,3%) e capoeira (10,9%). A causa para esta confusão, já mencionada, é a proximidade dos valores de intensidade média para as classes de formação naturais.

- Capoeirão (ME): esta formação natural foi corretamente classificada em 59,0%. A maior confusão é percebida com a floresta (19,9%). A floresta de galeria (CI) e capoeira (IN) foram confundidas respectivamente em 4,8% e 15,9%.

- Capoeira (IN): esta classe foi separada com um acerto de 54,6%. Como constatado anteriormente, a confusão entre as formações vegetais permanece. A maior confusão está relacionada com o capoeirinha (CS) com índice de 18,1%. O erro relacionado a classificação nas formações naturais floresta (AV) e capoeirão (ME) é de 11,7% e 5,9%. Para as áreas ocupadas por floresta de galeria (CI) o erro verificado foi de 9,5%.

- Reflorestamento (RE): as áreas ocupadas com *Pinnus sp.* foram corretamente classificadas em 87,7%. Percebe-se apenas confusão com a formação

natural bracatinga (BR). Neste caso, as áreas classificadas estão espacialmente próximas e com altitudes médias similares, o que levou o classificador a confundir as duas classes.

- Edificações (ED): esta classe antrópica foi a que apresentou pior desempenho na classificação com 19,3% de índice de acerto. A maior confusão (35,2%) ocorre com as áreas de agropecuária (AP). Conforme mencionado os telhados de amianto com resposta de intensidade média similar as áreas de agropecuária ocasionaram esta confusão. Com relação ao solo exposto (SO) o erro foi de 34,4%. A análise da base cartográfica permite visualizar que a maior parte das edificações (ED) está situada nas menores altitudes, justamente, o local onde está concentrada a maior área de solo exposto. Esta característica em comum ocasionou a confusão entre solo exposto e edificação.

- Várzea (VA): esta formação natural foi corretamente classificada em 60,2% dos casos. A proximidade dos valores de intensidade média associado ao fato destas duas classes estarem situadas em locais de altitude média baixa colaboraram para a confusão de 39,8% com a classe corpos d'água (AG).

- Corpos d'água (AG): o acerto para a classe água foi de 55,6% e conforme motivos já apresentados foi confundida em 42,3% com a formação natural várzea (VA).

- Solo exposto (SO): a classificação obteve um índice de acerto de 83,8% e conforme verificado a confusão com a agropecuária (AP) foi de 8,4% em virtude dos valores de intensidade média similares. Com as edificações (ED) a confusão resultou em 7,9% em função da similaridade de intensidade média dos telhados de amianto com o solo exposto.

Ao analisar as oitos classes relativas a cobertura vegetal (bracatinga, várzea, floresta de galeria, floresta, capoeirão, capoeira, reflorestamento e capoeirinha ) pode-se concluir:

- O melhor desempenho foi alcançado pelo reflorestamento (RE) 87,7%. A intensidade média associado ao fato de que a área ocupada por reflorestamento concentra-se em uma mesma faixa de altitude média, possibilitou uma boa separabilidade em relação às demais classes de formação vegetal.

- O pior desempenho ocorreu com a classe capoeira (IN) com 54,6%. Contudo as classes relativas a floresta (AV), capoeirão (ME) e floresta de galeria (CI) obtiveram desempenho similar.i

- O acerto médio ao considerar as formações vegetais foi de 65,1%. Considerando-se separadamente as quatro classes: solo exposto (SO), agropecuária (AP), corpos d'água (AG) e edificações (ED) o acerto médio obtido foi de 57,0%.

Com relação ao acerto médio para este primeiro experimento o valor obtido foi de 62,4%. A análise dos resultados do primeiro experimento fornecem subsídios para afirmar que o emprego das variáveis altitude e intensidade média não possibilitaram uma boa separabilidade entre as classes relativas as formações vegetais, com exceção a classe reflorestamento (RE).

#### 4.4.1.2 Grade regular com resolução de 10 metros empregando-se as variáveis H,I e altura Média na célula

Para o segundo experimento foi procedida a classificação introduzindo o cofator relativo a altura média da cobertura vegetal na célula (h). A motivação para a introdução desta variável é que ela é relacionada com o estágio das formações vegetais, sendo empregada na classificação destas formações em campo.

Com os valores provenientes das amostras para os três cofatores foi possível obter um espaço canônico igual a três e o ajuste forneceu três funções de classificação para este experimento:

$$\hat{Y}_1 = -73,808 + 0,089174.h - 0,00183.I + 0,07997.\Delta h \quad (125)$$

$$\hat{Y}_2 = 12,087 - 0,01848.h + 0,01949.I + 0,28910.\Delta h \quad (126)$$

$$\hat{Y}_3 = -1,326 - 0,00375.h + 0,05058.I - 0,06306.\Delta h \quad (127)$$

As equações (125), (126) e (127) possibilitaram o cálculo das coordenadas dos centróides relativos a cada população (classe temática). As coordenadas obtidas para os centróides são mostradas na Tabela 6.

TABELA 6 - CENTRÓIDES DAS CLASSES TEMÁTICA PARA A GRADE 10m - VARIÁVEIS  $h$ ,  $l$  e  $\Delta h$

Classe	$\hat{Y}_1$	$\hat{Y}_2$	$\hat{Y}_3$
AP	-3,51941	-0,26189	2,03447
AV	10,51140	-0,37637	-0,44193
BR	3,28935	-1,53033	1,22438
CS	-6,39470	0,13921	-0,44149
CI	-5,84045	-0,34617	-2,50196
ED	-8,07858	-1,82883	-1,18499
IN	-7,67982	-2,18273	-0,28813
ME	7,26310	0,03198	-0,04868
RE	5,47476	4,46683	-0,91570
AG	6,17810	-1,60223	0,41731
SO	9,00865	-2,24882	-0,63620
VA	-6,43569	1,75823	0,23538

A partir da classificação e da sobreposição com os resultados gerados pela fotogrametria, obteve-se a matriz de confusão que é apresentada na Tabela 7.

TABELA 7 - MATRIZ DE CONFUSÃO PARA A GRADE 10 m - VARIÁVEIS  $H$ ,  $I$  e  $\Delta h$

Classe	AP	BR	CS	CI	ED	AG	VA	AV	ME	IN	RE	SO	Soma
AP	0,665	0,000	0,148	0,000	0,021	0,000	0,000	0,000	0,000	0,000	0,032	0,134	1,000
BR	0,000	0,723	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,003	0,274	0,000	1,000
CS	0,088	0,000	0,683	0,032	0,000	0,006	0,000	0,016	0,023	0,123	0,002	0,026	1,000
CI	0,000	0,000	0,000	0,768	0,000	0,004	0,000	0,023	0,046	0,156	0,000	0,003	0,000
ED	0,427	0,000	0,006	0,000	0,186	0,000	0,000	0,003	0,000	0,000	0,000	0,378	1,000
AG	0,020	0,000	0,001	0,000	0,000	0,608	0,341	0,000	0,000	0,004	0,000	0,025	1,000
VA	0,000	0,000	0,000	0,017	0,004	0,377	0,599	0,000	0,000	0,000	0,000	0,004	1,000
AV	0,000	0,037	0,000	0,038	0,000	0,000	0,000	0,738	0,134	0,052	0,000	0,000	1,000
ME	0,000	0,000	0,000	0,012	0,000	0,001	0,002	0,156	0,716	0,097	0,016	0,000	0,000
IN	0,002	0,005	0,114	0,036	0,002	0,000	0,001	0,000	0,102	0,737	0,000	0,000	1,000
RE	0,000	0,082	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,918	0,000	1,000
SO	0,147	0,000	0,000	0,000	0,024	0,000	0,000	0,000	0,000	0,000	0,000	0,829	1,000

Para possibilitar uma comparação entre o primeiro e segundo experimento, elaborou-se a Tabela 8, na qual são apresentados os índices de acerto para as classes nestes experimentos.

TABELA 8 - ÍNDICES DE ACERTO PARA AS CLASSES NOS DOIS EXPERIMENTOS

CLASSE	1º EXPERIMENTO	2º EXPERIMENTO
AP	0,679	0,665
ED	0,193	0,186
AG	0,571	0,608
SO	0,838	0,829
BR	0,686	0,723
CS	0,636	0,683
CI	0,627	0,768
VA	0,602	0,599
AV	0,646	0,738
ME	0,589	0,716
IN	0,546	0,737
RE	0,877	0,918

Ao considerar os índices de acerto para a classificação do primeiro e do segundo experimento apresentados na tabela 8 obteve-se o gráfico mostrado na Figura 39.

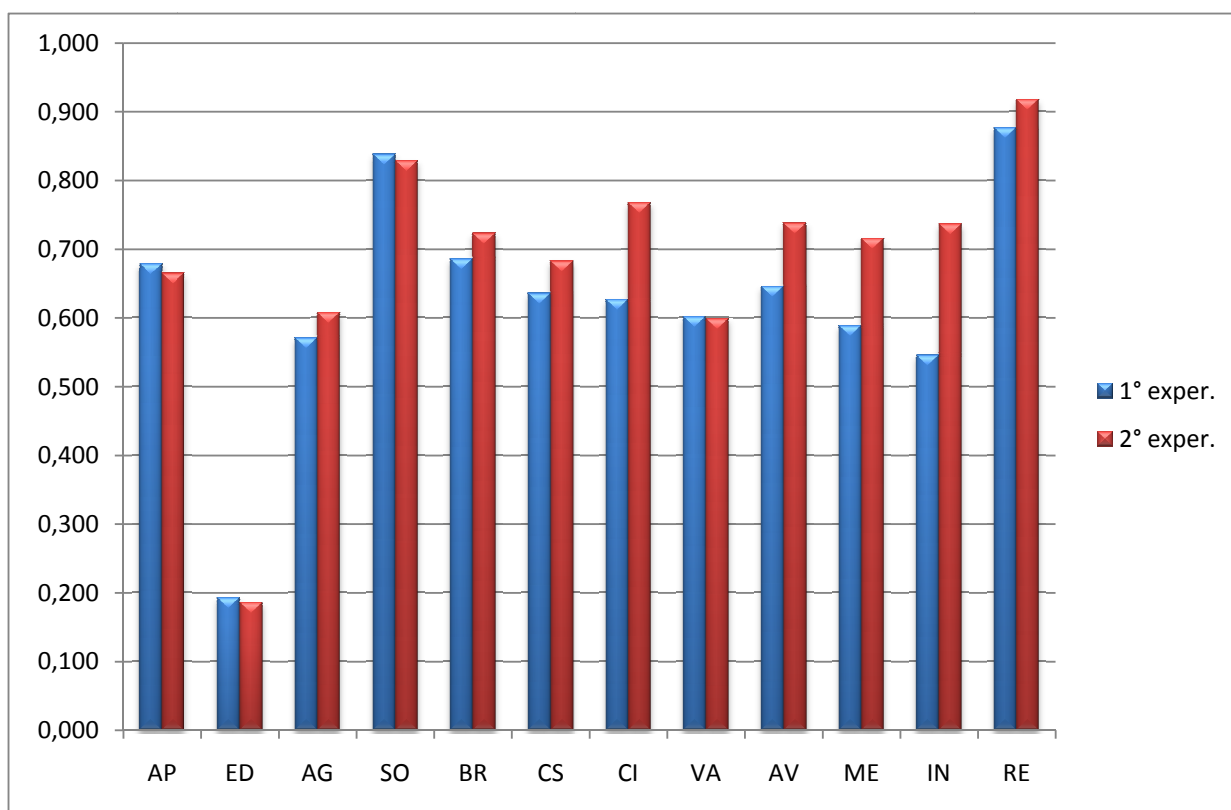


FIGURA 39 - ÍNDICES DE ACERTO DO 1º E 2º EXPERIMENTOS



Procedendo a análise da Tabela 7 e da Figura 39 pode-se concluir que:

- Agropecuária (AP): apresentou índice de acerto similar ao primeiro experimento, passando de 67,4% para 66,5%. A confusão com a classe capoeirinha (CS) de 10,5% aumentou para 14,8%. Este aumento está relacionado com a exatidão obtida na determinação do MDT na área de estudo. Os resultados alcançados com a filtragem matemática mostraram que a superfície ajustada apresentou uma tendência de -1,518 metro, ou seja, um rebaixamento em relação a superfície considerada como correta. Desta forma a altura média da célula, para classes próximos do solo como a agropecuária (AP) e solo (SO), apresentaram um valor acima de 1,518 m. O classificador identificou estas duas classes como elementos acima do solo e adiciona-se a isso, o fato exposto no primeiro experimento, no qual as áreas de capoeirinha (CS) são áreas agrícolas abandonadas.

A confusão com o solo exposto (SO) diminuiu em relação ao primeiro experimento passando de 16,6% para 13,4%. Para esta situação a altura média da célula não apresentou grande vantagem, pois as duas classes apresentam praticamente a mesma altura média na célula.

Os índices de confusão com a classes edificação (ED) e reflorestamento (RE) foram, respectivamente, de 2,1% e 3,2% não ocorrendo diferenças significativas em relação ao primeiro experimento. A confusão de 1,1% verificada no primeiro experimento para a classe várzea (VA) foi eliminada. Face ao exposto pode se afirmar que a inserção do terceiro cofator na classificação não melhorou o acerto para a classe agropecuária (AP).

- Bracatinga (BR): o acerto da classificação, no segundo experimento, da classe bracatinga (BR) foi de 72,3%, índice superior ao obtido no primeiro experimento de 68,6%. A confusão verificada com a classe reflorestamento (RE) baixou de 31,4% para 27,4%. Neste caso, a variável altura média na célula propiciou benefícios, o que pode ser comprovado pela diferença de altura entre as duas formações naturais. Na inspeção de campo verificou-se que a floresta de bracatinga (BR) é mais alta que a floresta ocupada com *Pinnus sp.*. O rebaixamento do MDT não apresentou influência nesta situação pois as duas formações naturais apresentam altura da cobertura vegetal acima de 12 metros.

- Capoeirinha (CS): nesta classe o acerto foi de 68,3%, valor superior ao obtido no primeiro experimento que foi de 63,6%. Ocorreu mistura com a classe

agropecuária (AP) em 8,8%, índice similar obtido no primeiro experimento (8,4%). A maior confusão ainda é percebida com a capoeira (IN) verificando-se um índice de 12,3%. Porém, este resultado é melhor que os 24,6% verificados no primeiro experimento. Neste caso, a melhora do índice está relacionada ao fato de que a capoeira (IN) apresenta árvores não observadas na capoeirinha (CS) e que fazem com que a altura média na célula seja maior. Os índices de confusão para solo exposto (SO), floresta de galeria (CI), floresta (AV) e capoeirão (ME) apresentaram valores de confusão pequenos, sendo respectivamente de 3,2%, 2,6%, 1,6% e 2,3%.

- Floresta de galeria (CI): esta cobertura do solo foi classificada corretamente com um índice de 76,8%, o que demonstra uma melhoria significativa quando comparado ao resultado de 62,7% obtidos no primeiro experimento. Esta melhoria de 14,1% no acerto está relacionada com a diminuição da confusão com o capoeirão (ME). Nesta situação o erro baixou de 19,5% para 4,6%. Devido a semelhança, da variável altura da cobertura vegetal, com a classe capoeira (IN), a confusão passou de 14,3% para 15,6%, portanto, não trazendo melhoria para este índice. A confusão com a classe floresta (AV) foi de 2,3%.

- Floresta (AV): neste caso, o índice de acerto alcançado pela classificador foi 73,8%, portanto, superior ao valor de 64,6% alcançado no primeiro experimento. Observa-se que a confusão com a classe floresta de galeria (CI) melhorou, passando de 9,2% para 3,8%. A mistura com a capoeira (IN) teve melhora de 5,7% baixando para 5,2%. Esta melhora nos índices reforça a importância da variável altura da cobertura vegetal na separação destes estágios de regeneração vegetal. A exceção foi com a classe capoeirão (ME) que apresentou um menor incremento no índice de confusão, passando de 15,3% para 13,4. A inspeção de campo possibilitou visualizar que estas duas coberturas vegetais apresentam alturas bastante próximas sendo esta a causa da confusão.

- Capoeirão (ME): no segundo experimento o índice de acerto desta classe foi de 71,6%, ou seja, um incremento de 7,6% quando comparado ao índice obtido no primeiro teste. A maior confusão ainda ocorre com a floresta (AV), verificando-se um índice de 15,6%, porém, menor que os 19,9% verificados no experimento anterior. A confusão com a floresta de galeria (CI) apresentou melhora passando de 4,8% para 1,2%. A mistura com a classe capoeira (IN) baixou de 15,9% para 9,7%.

Estes dois últimos resultados alcançados graças a utilização do cofator altura da cobertura vegetal.

- Capoeira (IN): o acerto obtido foi de 73,7% um aumento de 19,1% aos 54,6% obtidos com o primeiro experimento. A maior confusão ainda está relacionada com o capoeirinha (CS) com índice de 11,4%, porém, um valor 6,7% menor que os 18,1% obtidos anteriormente. A separação com a classe floresta (AV) foi alcançada totalmente com a configuração de cofatores utilizados. Porém, ocorreu um aumento na confusão com a classe capoeirão (ME) passando de 5,9% para 10,2%.

- Reflorestamento (RE): a classificação das áreas ocupadas por florestas plantadas mostrou melhoria de 4,1%, passando de 87,7% para 91,8%. A confusão que ocorria com a bracatinga (BR) aumentou de 4,9% para 8,2%. A causa para o aumento da confusão é que estas duas formações apresentam alturas similares.

- Edificações (ED): a classificação das edificações apresentou índice de acerto de 18,6%, um decréscimo no valor do primeiro experimento de 0,7%. Apesar do telhado das edificações estar posicionado acima do solo, deve ser destacado que a maior parte destas são casas de um pavimento, e portanto, o erro relacionado a determinação do MDT prejudicou o desempenho da altura da cobertura vegetal nesta situação. A maior confusão persiste com a classe agropecuária (AP), ocorrendo, neste caso, um aumento de 35,2% para 42,7%. O erro relacionado a classificação como solo exposto (SO) também aumentou passando de 34,4% para 37,8%. Para a classe edificação (ED) a inserção da variável altura da cobertura vegetal não auxiliou na melhoria do desempenho do classificador. Outro problema relacionado ao mau desempenho do classificador para as edificações está na resolução geométrica da grade que é de 10 metros. Das 209 edificações existentes na área de estudo, 133 apresentam área do telhado inferior a 70 m<sup>2</sup>, ou seja, menor que a área de uma célula que é de 100 m<sup>2</sup>.

- Várzea (VA): foi corretamente classificada em 59,9% dos casos, valor praticamente igual ao índice de 60,2% alcançado no primeiro experimento. A confusão com a classe água diminuiu, passando de 39,8% para 37,7%. O terceiro cofator utilizado no segundo experimento não possibilitou melhoria do índice de classificação desta classe.

- Corpos d'água (AG): o classificador obteve acerto de 60,8%, superior ao índice de 55,6% alcançado no primeiro experimento. A confusão com a várzea (VA) diminuiu para 34,1%, valor inferior aos 42,3% observados na primeira configuração

de variáveis. Assim como observado com a classe várzea (VA), a variável altura da cobertura vegetal não possibilitou uma melhor separação entre estas classes.

- Solo exposto (SO): esta classe teve seu desempenho degradado passando de um índice de acerto de 83,8% no primeiro experimento para 82,9%. A confusão com a agropecuária (AP) aumentou de 8,4% para 14,7%. A confusão com as edificações (ED) baixou de 7,9% para 2,4%. De forma geral a configuração de dados utilizada para o segundo experimento não trouxe benefícios para a classe solo exposto (SO).

Analisando os resultados das tabelas 7, 8 e o gráfico 38 para o grupo de classes relativas a cobertura vegetal pode-se afirmar que:

- O melhor desempenho foi alcançado pelo reflorestamento (RE) 91,8%. Esta classe já havia conseguido um bom resultado no primeiro experimento e o acréscimo da variável altura da cobertura vegetal propiciou a melhoria no índice de acerto.

- O pior desempenho foi verificado para a classe várzea (VA) 59,9%. Nesta classe a altura da cobertura vegetal ajudou a piorar o índice de acerto.

- O maior acréscimo no índice de acerto é percebido para a capoeira passando de 54,6% para 72,3%.

- O acerto médio considerando-se somente as formações vegetais foi de 73,5%. A partir deste valor pode ser afirmado que a altura da cobertura vegetal ocasionou a melhoria refletida na menor confusão entre as formações naturais.

- Para as quatro classes: solo exposto (SO), agropecuária (AP), água (AG) e edificações (ED) o acerto médio obtido foi de 57,2% praticamente igual ao índice de 57,0% obtido no primeiro experimento. Para este grupo de classes a configuração de cofatores empregada no segundo experimento não propiciou melhorias nos índices de acerto.

O acerto médio do segundo experimento foi de 68,1%, superior ao índice de 62,4% alcançado com a primeira configuração de cofatores. A adição da variável cobertura vegetal propiciou a melhora observada no acerto médio, possibilitando uma melhor separação entre as formações naturais em seus diferentes estágios de sucessão.

#### 4.4.1.3 Grade regular com resolução de 10 metros empregando-se todas as variáveis

Com o objetivo de melhorar os índices de acerto relativos ao grupo de classes antrópica e também visando uma maior separabilidade entre as classes formação natural em estágio avançado e médio de regeneração implementou-se o terceiro experimento. Nesta situação empregaram-se todas as variáveis: Altitude Média(h), Intensidade Média(I), Altura média da cobertura vegetal ( $\Delta h$ ), Desvio Padrão da Altitude( $\sigma h$ ), Desvio Padrão da Intensidade ( $\sigma I$ ) e Densidade de Pontos (De).

A partir das amostras ajustou-se as 6 equações de discriminação, cujos coeficientes são mostrados na Tabela 9.

TABELA 9 - COEFICIENTES DAS FUNÇÕES DE DISCRIMINAÇÃO PARA A GRADE REGULAR DE 10 m EMPREGANDO TODAS AS VARIÁVEIS

Função	h	I	$\Delta h$	$\sigma H$	$\sigma I$	De	Constante
$\hat{Y}_1$	0,09277	0,00115	0,04537	-0,11933	0,00996	0,87532	-78,35670
$\hat{Y}_2$	-0,01576	0,01137	0,29000	0,22009	0,02284	-1,21332	11,65110
$\hat{Y}_3$	-0,01865	0,02845	-0,06903	0,44061	0,10804	2,27860	6,06356
$\hat{Y}_4$	0,00179	0,04524	-0,03173	0,19485	-0,08249	-0,40956	-3,19649
$\hat{Y}_5$	0,00000	0,00115	-0,05719	-0,91753	0,16001	-1,29507	-0,18246
$\hat{Y}_6$	0,00572	-0,01427	-0,08636	0,78176	-0,01095	-1,07782	-2,21513

Com ajuste das funções de discriminação calcularam-se as coordenadas dos centróides relativos a cada população (classe temática). As coordenadas obtidas para os centróides no terceiro experimento são mostradas na Tabela 10.

TABELA 10 - CENTRÓIDES DAS CLASSES TEMÁTICA PARA A GRADE 10m - TODAS AS VARIÁVEIS

Classe	$\hat{Y}_1$	$\hat{Y}_2$	$\hat{Y}_3$	$\hat{Y}_4$	$\hat{Y}_5$	$\hat{Y}_6$
AP	-3,42498	-0,80913	0,88929	2,02433	-0,37251	-0,37341
AV	10,75150	-0,17329	-2,66620	0,32220	0,00756	-0,19045
BR	3,36430	-1,28673	-0,68458	1,52684	0,16304	0,69550
CS	-6,76247	0,31545	1,05419	-0,55401	-0,64601	0,77446
CI	-5,65570	-1,42564	-0,89447	-1,79452	-2,39263	-0,73118
ED	-8,23835	-2,19906	-1,18983	-0,91949	-0,08574	0,01808
IN	-7,91303	-2,09494	-0,52705	-0,58281	1,44023	0,37554
ME	7,77802	0,18895	1,51528	-0,44773	-0,18981	0,38589
RE	4,71802	5,49879	-0,78067	-0,73221	0,12746	-0,15902
AG	7,59669	-2,54181	2,79672	-0,77978	0,83872	-0,99625
SO	10,19400	-2,85309	0,43167	-0,88563	-0,05350	-0,33469
VA	-7,08621	1,96852	0,24668	0,14427	0,19934	-0,15125

Com o ajuste das funções discriminantes e o cálculo dos centróides das classes procedeu-se a classificação da grade regular de 10 m. A classificação obtida foi confrontada com a base cartográfica obtida pela restituição fotogramétrica e então se obteve a matriz de confusão que é apresentada na Tabela 11.

TABELA 11 - MATRIZ DE CONFUSÃO PARA A GRADE 10 m - TODAS AS VARIÁVEIS

Classe	AP	BR	CS	CI	ED	AG	VA	AV	ME	IN	RE	SO	Soma Parcial
AP	0,834	0,000	0,023	0,000	0,053	0,000	0,000	0,000	0,000	0,002	0,000	0,088	1,000
BR	0,000	0,855	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,017	0,128	0,000	1,000
CS	0,020	0,008	0,701	0,000	0,026	0,006	0,003	0,014	0,055	0,148	0,009	0,011	1,000
CI	0,003	0,000	0,000	0,703	0,000	0,000	0,004	0,028	0,015	0,247	0,000	0,000	1,000
ED	0,353	0,000	0,002	0,000	0,306	0,005	0,005	0,006	0,000	0,000	0,000	0,324	1,000
AG	0,006	0,000	0,002	0,006	0,001	0,625	0,310	0,000	0,000	0,033	0,000	0,018	1,000
VA	0,004	0,000	0,000	0,014	0,000	0,343	0,638	0,000	0,000	0,000	0,000	0,001	1,000
AV	0,000	0,032	0,000	0,001	0,006	0,000	0,001	0,768	0,120	0,068	0,003	0,000	1,000
ME	0,000	0,009	0,000	0,001	0,002	0,001	0,003	0,093	0,792	0,072	0,026	0,000	1,000
IN	0,005	0,004	0,069	0,075	0,000	0,000	0,010	0,034	0,045	0,754	0,003	0,001	1,000
RE	0,000	0,098	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,000	0,902	0,000	1,000
SO	0,107	0,000	0,000	0,003	0,015	0,002	0,005	0,000	0,000	0,001	0,000	0,867	1,000

Procedendo de maneira similar a comparação realizada na seção 4.1.1.2 foi obtida a tabela comparando os índices de acerto para as classes no primeiro,

segundo e terceiro experimentos. Esta comparação pode ser verificada na Tabela 12.

TABELA 12 - ÍNDICES DE ACERTO DAS CLASSES PARA OS TRÊS EXPERIMENTOS

CLASSE	1º EXPERIMENTO	2º EXPERIMENTO	3º EXPERIMENTO
AP	0,679	0,665	0,834
ED	0,193	0,186	0,306
AG	0,571	0,608	0,625
SO	0,838	0,829	0,867
BR	0,686	0,723	0,855
CS	0,636	0,683	0,701
CI	0,627	0,768	0,703
VA	0,602	0,599	0,638
AV	0,646	0,738	0,768
ME	0,589	0,716	0,792
IN	0,546	0,737	0,754
RE	0,877	0,918	0,902

A partir dos valores apresentados na tabela 12 criou-se o gráfico apresentado na Figura 40.

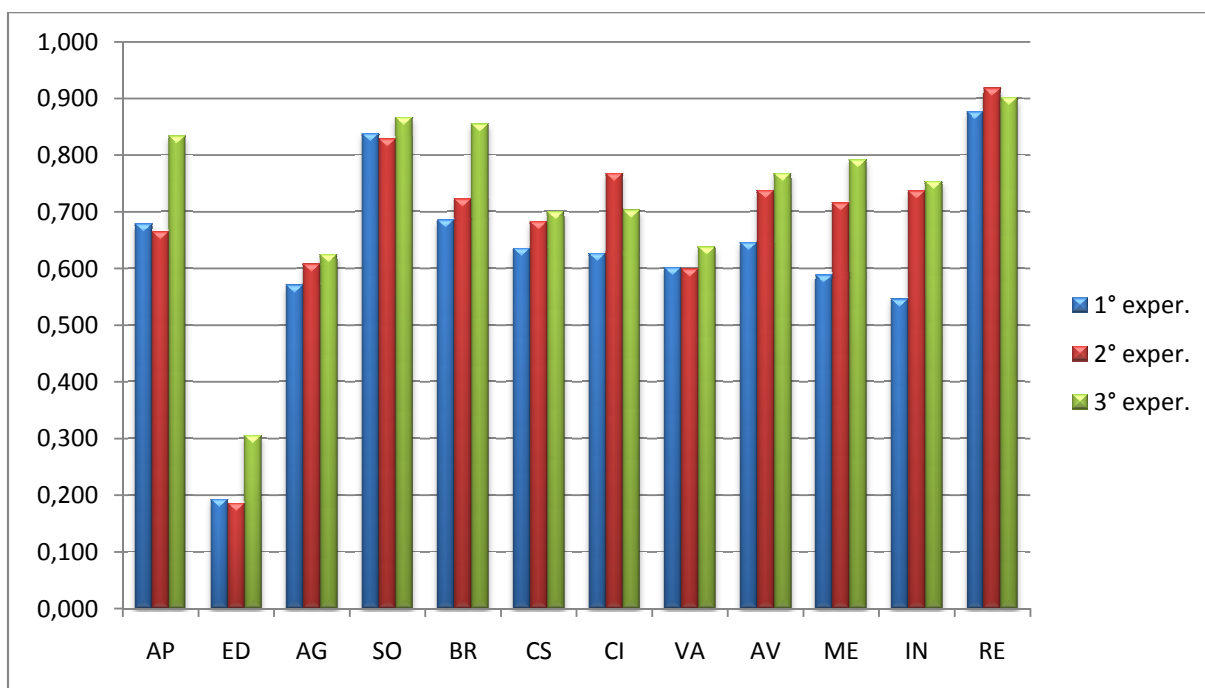


FIGURA 40 - ÍNDICES DE ACERTO DO 1º, 2º E 3º EXPERIMENTO

A comparação do segundo com o terceiro experimento permite afirmar que:

- Agropecuária (AP): esta classe apresentou melhoria no índice de acerto comparativamente ao segundo experimento, passando de 66,5% para 83,4%. A melhoria deste valor está diretamente ligada a diminuição da confusão com a classe capoeirinha (CS). No segundo experimento o valor de confusão foi de 14,8%, baixando para 2,3%. A melhoria desta condição está relacionada diretamente com a inserção da variável densidade de pontos. Nas áreas ocupadas por agropecuária (AP) o pulso LASER reflete diretamente no solo, fazendo com que ocorra basicamente um retorno. Nas áreas de capoeirinha (CS) ocorrem formações vegetais de baixa estatura e que ocasionam conforme a incidência do pulso LASER mais de um retorno, fazendo com que a densidade de pontos seja maior.

A confusão com o solo exposto (SO) melhorou em relação ao segundo experimento passando de 13,4% para 8,8%. O cofator desvio padrão da intensidade tem relação direta com esta melhoria. Nas áreas de solo exposto (SO) a intensidade média é mais uniforme e, portanto, apresentando valores de desvio padrão da intensidade menores.

O índice de confusão com a classe edificação (ED) foi de 5,3%, maior quando comparado ao segundo experimento (2,1%).

- Bracatinga (BR): o índice de acerto obtido no terceiro experimento para esta classe foi de 85,5%, índice 13,2% superior ao valor obtido no segundo teste (72,3%). A melhora neste índice está relacionada com o decréscimo na confusão com a classe reflorestamento (RE) que passou de 27,4% para 12,8%. A causa para esta diminuição é que as áreas ocupadas por bracatinga (BR) são mais abertas que as ocupadas com *Pinnus sp.*, apresentando uma densidade de pontos maior que auxiliou o desempenho da FDL empregando todas as variáveis. Para esta classe a configuração de cofatores utilizados para a classificação trouxe grandes benefícios.

- Capoeirinha (CS) : o acerto desta classe foi de 70,1% superior aos 68,3% determinados no experimento anterior. A configuração de variáveis utilizadas neste teste não trouxe benefícios para esta classe pois a confusão com a capoeira (IN) aumentou de 12,3% para 14,8%. A mistura com as classes capoeirão (ME) e floresta (AV) permanece, confirmando nesta situação os problemas relacionados a geração do MDT.

- Floresta de galeria (CI): esta classe foi classificada corretamente em 70,3% dos casos, o que representa uma deterioração do índice de 76,8% obtido no segundo experimento. Esta piora está relacionada com as similaridades nos valores



de densidade e na altura da cobertura vegetal quando confrontada com a classe capoeira (IN). Nesta comparação a confusão aumentou de 15,6%, no segundo teste, para 24,7% no terceiro teste.

A confusão com floresta (AV) e capoeirão (ME) foi respectivamente de 2,8% e 1,5%.

- Floresta (AV): na configuração de variáveis empregadas o índice de acerto alcançado pelo classificador para esta classe foi de 76,8%, uma melhoria de 3% relação ao valor do segundo teste. A confusão com a capoeirão (ME) praticamente não sofreu alterações baixando de 13,4% para 12,0%. A confusão com a capoeira (IN) permanece aumentando de 5,2% para 6,8%. A confusão entre os estágios de regeneração da vegetação está diretamente ligada à qualidade do MDT e, portanto, o acréscimo dos cofatores densidade, desvio padrão da altitude e de intensidade não acrescentou melhora na separação destas classes.

- Capoeirão (ME): o índice de acerto obtido foi de 79,2%, ou seja, um aumento de 7,6% em relação ao segundo experimento. A confusão com a floresta (AV) permanece sendo obtido o valor de 9,3%, reflexo da qualidade do MDT.

Para este experimento não ocorreu confusão do capoeirão (ME) com a floresta de galeria (CI). A causa para esta melhora pode ser observada no desvio padrão da altitude, que nas classes floresta (AV) e capoeirão (ME) é maior que o observado na floresta de galeria (CI). Esta classe apresenta altura da cobertura vegetal menor e copas menores que fazem com que a diferença de altitude do primeiro e último retorno LASER seja baixa. A densidade também está relacionada com esta melhora pois seu valor é menor na floresta de galeria (CI).

A mistura com a classe capoeira (IN) baixou de 9,7% para 7,2%. Não trazendo melhora significativa para o índice de confusão.

- Capoeira (IN): o acerto obtido foi de 75,4%, valor similar ao observado no segundo experimento que foi de 73,7%. A confusão com o capoeirinha (CS) apresentou melhora passando de 11,4% para 6,9%. A causa para esta melhora está na densidade de pontos LASER que é maior na capoeira (IN), sendo isto reflexo da maior quantidade de indivíduos arbóreos presentes nesta formação. A confusão com as áreas de floresta (AV) e capoeirão (ME) foi respectivamente de 3,4% e 4,5%, um resultado similar ao obtido no segundo experimento.

- Reflorestamento (RE): as áreas ocupadas com *Pinnus sp.* foram corretamente classificadas com um índice de 90,2%, não ocorrendo diferença

significativa com o índice alcançado no segundo experimento que foi de 82,9%. A confusão com a formação natural bracatinga (BR) também não apresentou melhoria sendo obtidos 9,8% contra 8,2% verificados no segundo teste. Portanto, o emprego das seis variáveis não possibilitou a melhoria na classificação das áreas de reflorestamento (RE).

- Edificações (ED): a classificação das edificações apresentou aumento de 12% no índice de acerto passando de 18,6% para 30,6% no terceiro teste. A confusão com a classe agropecuária (AP) diminuiu passando de 42,7% para 35,3% um índice alto de erro. A mistura com a classe diminuiu de 37,8% para 32,4%. A configuração de variáveis empregada não foi suficiente para a melhoria do acerto do classificador no caso das edificações (ED). A resolução geométrica da grade e a qualidade do MDT tem relação direta com esta condição.

- Várzea (VA): foi corretamente classificada em 63,8% dos casos. Houve uma melhoria de 3,9% quando comparado com o índice alcançado no segundo experimento. A confusão com a classe corpos d'água (AG) diminuiu, passando de 37,7% para 34,3%. Assim como pode ser verificado com a classe temática corpos d'água (AG) os cofatores adicionados apresentam similaridade com a classe várzea (VA), e portanto, não ofereceram melhoria significativa na separação destas duas classes.

- Corpos d'água (AG): o classificador obteve acerto de 62,5% que é superior ao valor de 60,8% obtido no segundo teste. A confusão com a várzea (VA) diminuiu de 34,1% para 31,0%, não sendo perceptível melhoria na classificação.

- Solo exposto (SO): o desempenho desta classe melhorou passando de 82,9% para 86,7%. A confusão com a agropecuária (AP) diminuiu passando de 14,7% para 10,7%. A confusão com as edificações (ED) permaneceu baixa passando de 2,4% para 1,5%. A configuração do terceiro teste não apresentou melhorias para a classificação do solo exposto (SO).

Analisando os resultados das Tabelas 11,12 e do gráfico 39 para o grupo de classes relativas a cobertura vegetal pode-se afirmar que:

- O melhor desempenho foi alcançado na classificação da classe reflorestamento (RE), com 90,2% de acerto. A configuração de variáveis não ocasionou melhoria na classificação desta formação vegetal.

- O menor desempenho do classificador pode ser verificado no acerto da classe temática várzea (VA). Quando comparada com o segundo experimento apresentou pequena melhoria passando de 59,9% para 63,8%.

- Destaca-se ainda o caso da floresta de galeria (CI) cujo índice de acerto baixou de 76,8% para 70,3%.

- O maior acréscimo no índice de acerto é percebido para a classe bracinga (BR) cujo índice de acerto passou de 72,3% para 85,5%. Portanto, no terceiro experimento ocorreu um acréscimo de 13,2%.

- O acerto médio considerando-se somente as formações vegetais foi de 76,4%, superior portanto ao índice de 73,5% obtido no segundo teste.

- Para as quatro classes: solo exposto (SO), agropecuária (AP), corpos d'água (AG) e edificações (ED) o acerto médio obtido foi de 65,8% superior ao índice médio de 57,2% obtido no segundo experimento. Para este grupo de classes a configuração de cofatores empregada no segundo experimento propiciou melhorias significativas principalmente para a classe agropecuária (AP) com um aumento de 16,9% no índice de acerto. Inclusive, foi o melhor incremento obtido para o terceiro experimento.

O acerto médio do terceiro experimento foi de 72,9%, portanto superior aos 68,1% obtidos com a segunda configuração de cofatores. Verifica-se com os resultados alcançados no terceiro experimento que a utilização de todos os cofatores constitui-se na melhor opção para a classificação dos dados LASER.

#### 4.4.1.4 Grade regular com resolução de 5 metros empregando-se todas as variáveis

Este experimento empregando uma resolução de 5 metros tem por objetivos verificar se ocorrem melhorias na classificação com o aumento da resolução geométrica da grade regular.

Para esta classificação obteve-se o ajuste das 6 equações discriminantes para a grade de 5 metros. A Tabela 13 apresenta os coeficientes obtidos para as funções de discriminação.

TABELA13 - COEFICIENTES DAS FUNÇÕES DE DISCRIMINAÇÃO PARA A GRADE REGULAR DE 5 m EMPREGANDO TODAS AS VARIÁVEIS

Função	h	l	$\Delta h$	$\sigma H$	$\sigma l$	De	Constante
$\hat{Y}_1$	0,09517	-0,00045	0,04044	-0,11520	-0,00306	0,58692	-79,39720
$\hat{Y}_2$	-0,01636	0,01537	0,27820	0,14370	0,02704	-0,84156	11,33270
$\hat{Y}_3$	-0,01549	0,02531	-0,05615	0,28150	0,09682	1,91243	5,12531
$\hat{Y}_4$	0,00101	0,04147	-0,04346	0,12575	-0,04836	-0,44383	-2,78418
$\hat{Y}_5$	0,00006	-0,00118	-0,05598	-0,85523	0,13091	-1,11320	0,20798
$\hat{Y}_6$	0,00430	-0,00768	-0,07610	0,77341	-0,00439	-0,98427	-1,73517

A partir das funções discriminantes cujos coeficientes são apresentados na Tabela 13 foram calculados os centróides de cada classe. Estes centróides são mostrados na Tabela 14.

TABELA 14 - CENTRÓIDES DAS CLASSES TEMÁTICAS PARA A GRADE 5m - TODAS AS VARIÁVEIS

Classe	$\hat{Y}_1$	$\hat{Y}_2$	$\hat{Y}_3$	$\hat{Y}_4$	$\hat{Y}_5$	$\hat{Y}_6$
AP	-3,51872	-0,59414	0,72148	1,72351	-0,29491	-0,25560
AV	11,01960	-0,15505	-1,95959	0,19382	-0,09764	-0,10500
BR	3,60663	-1,45426	-0,91587	1,39171	0,18013	0,37777
CS	-6,64788	0,19968	0,98704	-0,72587	-0,53338	0,53760
CI	-5,71409	-1,36140	-0,79297	-1,77556	-1,93381	-0,95194
ED	-8,16557	-2,08152	-0,85483	-0,77037	0,09463	-0,03276
IN	-7,81515	-2,24200	-0,73676	-0,41989	0,95956	0,40721
ME	7,85430	0,40244	1,10612	-0,32335	-0,10270	0,39154
RE	4,62414	5,03428	-0,69866	-0,73706	0,29333	-0,33014
AG	7,45049	-2,34740	2,58657	-0,44138	0,66950	-0,79063
SO	10,16900	-2,43091	0,40650	-0,68613	-0,20996	-0,16019
VA	-7,09496	1,78174	0,14355	0,20356	0,09184	-0,04813

O próximo passo foi classificar cada célula da grade de 5 metros. Concluída a classificação procedeu-se a obtenção da matriz de confusão apresentada na Tabela 15.

TABELA 15 - MATRIZ DE CONFUSÃO PARA A GRADE 5 m - TODAS AS VARIÁVEIS

Classe	AP	BR	CS	CI	ED	AG	VA	AV	ME	IN	RE	SO	Soma
AP	0,659	0,000	0,098	0,005	0,073	0,000	0,000	0,008	0,004	0,002	0,006	0,145	1,000
BR	0,000	0,732	0,000	0,000	0,000	0,000	0,000	0,009	0,000	0,012	0,247	0,000	1,000
CS	0,038	0,002	0,635	0,008	0,063	0,002	0,001	0,012	0,036	0,165	0,014	0,023	1,000
CI	0,008	0,000	0,000	0,576	0,007	0,000	0,011	0,033	0,065	0,297	0,000	0,003	1,000
ED	0,284	0,000	0,000	0,000	0,395	0,024	0,031	0,004	0,000	0,006	0,001	0,254	1,000
AG	0,002	0,000	0,003	0,000	0,036	0,524	0,394	0,000	0,000	0,009	0,000	0,032	1,000
VA	0,000	0,000	0,000	0,019	0,017	0,366	0,588	0,000	0,000	0,000	0,000	0,010	1,000
AV	0,012	0,040	0,003	0,010	0,007	0,000	0,001	0,654	0,162	0,097	0,014	0,000	1,000
ME	0,014	0,018	0,011	0,002	0,001	0,002	0,002	0,143	0,704	0,075	0,028	0,000	1,000
IN	0,007	0,018	0,122	0,197	0,006	0,004	0,002	0,067	0,077	0,498	0,000	0,002	1,000
RE	0,000	0,104	0,000	0,000	0,000	0,000	0,000	0,002	0,000	0,009	0,885	0,000	1,000
SO	0,146	0,000	0,000	0,006	0,078	0,018	0,027	0,000	0,000	0,000	0,000	0,725	1,000

Também se realizou a comparação dos índices de acerto para as classes temáticas do terceiro e quarto experimentos. A Tabela 16 apresenta esta comparação.

TABELA 16 - ÍNDICES DE ACERTO PARA AS CLASSES NO 3° E 4° EXPERIMENTO

CLASSE	3°EXPERIMENTO	4°EXPERIMENTO	INCREMENTO
AP	0,834	0,659	-0,175
ED	0,306	0,395	0,089
AG	0,625	0,524	-0,101
SO	0,867	0,725	-0,142
BR	0,855	0,732	-0,123
CS	0,701	0,635	-0,066
CI	0,703	0,576	-0,127
VA	0,638	0,588	-0,050
AV	0,768	0,654	-0,114
ME	0,792	0,704	-0,088
IN	0,754	0,498	-0,256
RE	0,902	0,885	-0,017

A partir da Tabela 16 obteve-se o gráfico mostrado na Figura 41.

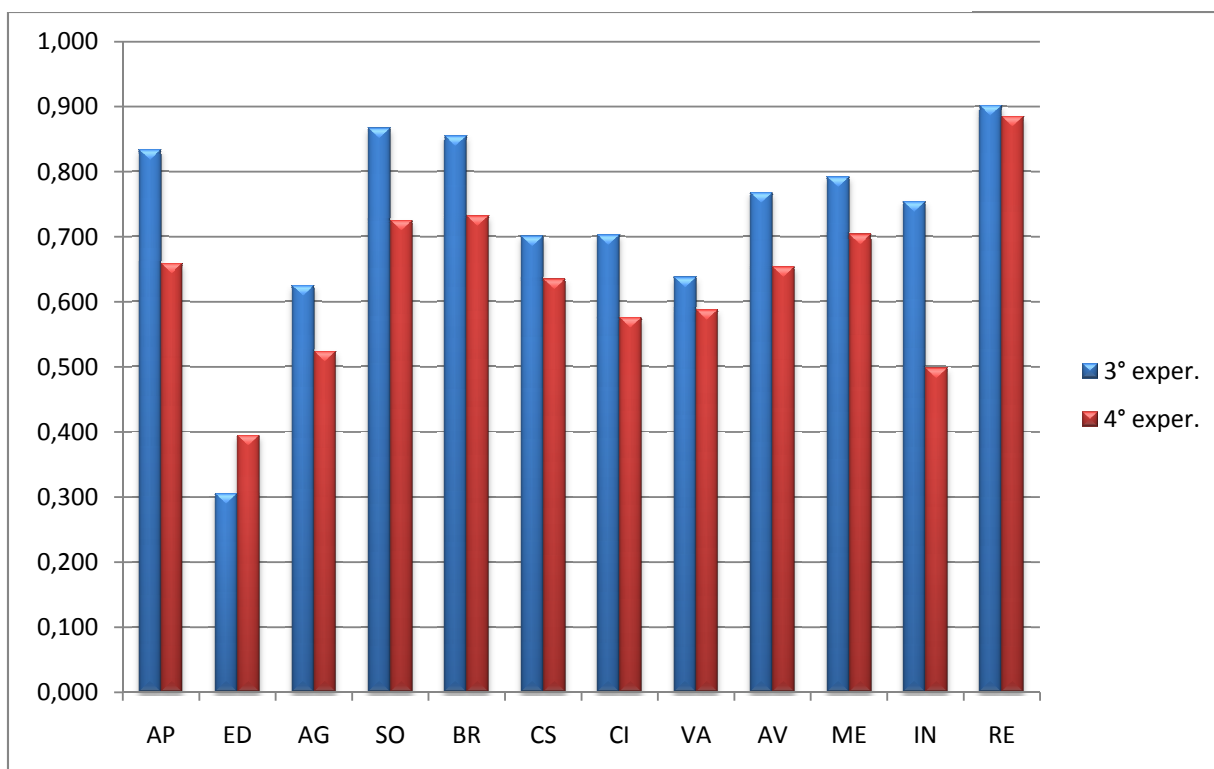


FIGURA 41 - ÍNDICES DE ACERTO DO 3º E DO 4º EXPERIMENTO

A análise Tabelas 11, 15 e 16 além do gráfico apresentado na Figura 41 auxiliam na elaboração das seguintes discussões:

- Agropecuária (AP): o índice de acerto de 83,4% obtido no terceiro experimento baixou para 65,9%. A confusão com a classe capoeirinha (CS) aumentou de 2,3% para 9,8%. A causa para a deterioração de 17,5% está no aumento da mistura com as áreas de capoeirinha (CS) e solo exposto (SO). A causa disto é o aumento da resolução espacial da grade regular, que possibilitou um detalhamento para o interior das áreas analisadas e que não era tão sensível anteriormente devido a maior generalização provocada pela resolução geométrica de 10 metros. A Figura 42 apresenta a imagem de intensidade derivada a partir do cofator intensidade e ilustrando o detalhamento provocado pelo aumento da resolução da grade regular de 10 metros (a), 5 metros (b) e 1 metro (c).

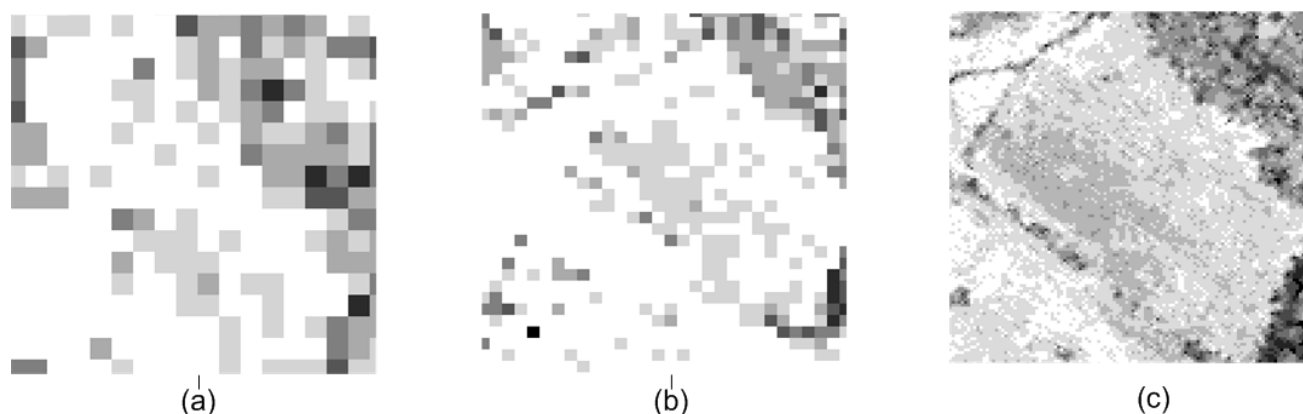


FIGURA 42 - EFEITO DO AUMENTO DA RESOLUÇÃO GEOMÉTRICA DA GRADE REGULAR

A confusão com o solo exposto (SO) e capoeirinha (CS) aumentou, respectivamente, de 8,8% para 14,5% e de 2,3% para 9,8%. O motivo é a proximidade de valores de intensidade média e também pelo maior detalhamento no interior das áreas.

Com a classe edificação (ED) também ocorreu um aumento na confusão, passando de 5,3% para 7,3%.

- Bracatinga (BR): no quarto experimento o acerto da bracatinga diminuiu de 85,5% para 73,2%, ou seja, de 12,3% do terceiro para o quarto experimento. A causa para este decréscimo é diretamente relacionada ao aumento na confusão com as áreas de reflorestamento (RE) que aumentou de 12,8% para 24,7%. A maior penetração do pulso LASER nas áreas de bracatinga é mais evidente na grade com maior resolução geométrica ocasionando resposta de intensidade no solo e nas copas. Este fenômeno fez com que a resposta da intensidade média da bracatinga (BR) se aproximasse dos valores observados nas áreas de *Pinnus sp.*.

- Capoeirinha (CS): o índice de 70,1% obtido no experimento anterior deteriorou para 63,5%, um decréscimo de 6,6%. Por se tratar de uma área agrícola abandonada a mistura com a agropecuária (AP) ocorreu com um valor de 3,8% e também se verifica a mistura com o solo exposto (SO) em 2,3% dos casos. A maior confusão persiste com as áreas de capoeira (IN) com 16,5%, um valor maior que a confusão obtida no terceiro experimento que foi de 14,8%. Deve-se lembrar que o capoeirinha (CS) é um estágio de regeneração anterior a classe capoeira.

- Floresta de galeria (CI): o aumento da resolução geométrica da grade fez com que o índice de acerto desta classe baixasse de 70,3% para 57,6%, uma perda de 12,7%. A mistura com a capoeira (IN) aumentou de 24,7% para 29,7% devido as similaridades verificadas entre estas duas formações vegetais relacionadas a altura média da floresta e também a resposta de intensidade.

- Floresta (AV): a configuração empregada neste experimento fez com que o índice de acerto de 76,8% diminuísse para 65,4% neste teste. A maior confusão ainda é verificada com a capoeirão (ME) que passou de 12,0% para 16,2%.

- Capoeirão (ME): o índice de acerto obtido de 79,2%, baixou para 70,4%. A confusão com a floresta (AV) aumentou de 9,3% para 14,3%. A qualidade do MDT ficou mais evidente na grade regular de 5 metros e isto causou o aumento desta confusão. A mistura com a classe capoeira (IN) aumentou de 7,2% para 7,5%.

- Capoeira (IN): esta foi a classe temática mais penalizada com o aumento da resolução geométrica da grade. Seu índice de acerto sofreu um decréscimo de 25,6% passando de 75,4% obtidos no terceiro experimento para 49,8%. A confusão aumentou significativamente com o capoeirinha (CS) passando de 6,9% para 12,2% devido de esta classe ser um estágio anterior na sucessão da vegetação. Com a floresta de galeria (CI) o erro passou de 7,5% para 19,7%. A similaridades entre estas duas formações foi responsável por este resultado.

- Reflorestamento (RE): a classificação das áreas de florestas plantadas de *Pinnus sp.* não sofreram redução significativa no índice de acerto. Este valor baixou de 90,2% para 88,5%. As características peculiares desta formação natural a preservaram dos efeitos do aumento da resolução geométrica da grade regular. A única classe similar ao reflorestamento (RE) e que gerou confusão foi a bracatinga (BR) com um índice de 10,4%.

- Edificações (ED): a única classe temática beneficiada com a configuração adotada no quarto experimento foi a edificação. Neste caso, o índice de acerto passou de 30,6% para 39,5%, ou seja, um aumento de 8,9%. A maior definição das bordas das edificações auxiliou nesta melhora. Deve ser recordado que das 209 edificações na área de estudo, 133 apresentam área do telhado inferior a 70 m<sup>2</sup>. Com a resolução geométrica de 10 metros a área mínima era de 100 m<sup>2</sup>. Ao empregar a grade com 5 metros de resolução geométrica este valor baixou para 25 m<sup>2</sup>. Este índice só não foi melhor porque ainda ocorreu confusão com as classes agropecuária (AP) e solo exposto (SO) que juntas totalizam 53,8% do erro.



- Várzea (VA): o acerto para a várzea baixou de 63,8% para 58,8%, uma redução de 5%. A maior confusão ainda ocorre com a classe corpos d'água (AG) com um erro de 36,6%, valor superior ao verificado no terceiro teste que foi de 34,3%.

- Corpos d'água (AG): o classificador obteve acerto de 52,4% valor inferior aos 62,5% verificados no terceiro experimento. A maior confusão é com a classe várzea (VA) com valor de 39,4% superior ao índice anterior de 31,0%.

- Solo exposto (SO): o desempenho do classificador (72,5%) para esta classe foi inferior ao do terceiro experimento (86,7%). A classe solo exposto teve seu acerto penalizado em 14,2% . A maior causa para esta depreciação está na confusão com a classe agropecuária (AP) com um índice de 14,5% por razões já comentadas.

Considerando-se o desempenho dos dois grupos (antrópico e formação natural) pode-se afirmar que:

- Para as formações naturais o aumento da resolução geométrica da grade não trouxe benefício ao processo de classificação. O índice de acerto médio para este grupo baixou de 76,4% para 65,9%. Neste grupo a classe mais penalizada foi a capoeira (IN) com um decréscimo de 25,6%, devido a maior confusão verificada com a floresta de galeria (CI) que apresenta estrutura florestal similar. A classe temática capoeirinha (CS) foi a menos penalizada com um decréscimo de 6,6% em seu índice de acerto.

- Para o grupo composto pelas classes edificação (ED), solo exposto (SO), agropecuária (AP) e corpos d'água (AG) o índice médio de acerto baixou de 65,8% do terceiro experimento para 57,6% no quarto teste. A classe mais penalizada foi a agropecuária (AP) com uma perda de 17,5% no seu índice de acerto. A classe edificação (ED) foi a única beneficiada com um aumento de 8,9% no acerto do classificador.

O experimento apresentou acerto médio de 69,2%, valor inferior ao conseguindo no terceiro teste que foi de 72,9%. De maneira geral pode-se afirmar que para a configuração de dados utilizada para a classificação e com densidade média de pontos próxima de 2 metros, obtida na varredura LASER, o aumento da resolução geométrica para 5 metros ocasionou mais perdas devido ao maior variabilidade dos parâmetros no interior dos limites das classes temáticas (conforme pode ser visualizado na Figura 42). A qualidade do MDT também foi determinante

nesta piora, pois o erro na sua determinação foi mais facilmente detectado pelo menor tamanho da célula da grade regular.

O quinto experimento empregando-se a grade de 1 metro foi executado visando principalmente atestar os problemas encontrados com o quarto experimento. A expectativa era que a confusão verificada aumentasse significativamente.

#### 4.4.1.5 Grade regular com resolução de 1 metro empregando-se todas as variáveis

De forma similar a seção anterior esta classificação foi obtida para a grade com resolução de 1 metro.

O processo de ajuste forneceu 6 equações discriminantes para a grade de 1 metro. A Tabela 17 apresenta os coeficientes obtidos para as funções de discriminação.

TABELA 17 - COEFICIENTES DAS FUNÇÕES DE DISCRIMINAÇÃO PARA A GRADE REGULAR DE 1 m EMPREGANDO TODAS AS VARIÁVEIS

Função	h	l	$\Delta h$	$\sigma H$	$\sigma l$	De	Constante
$\hat{Y}_1$	0,09370	-0,00360	0,02982	-0,15733	0,00930	0,89199	-79,55700
$\hat{Y}_2$	-0,01665	0,00941	0,23745	0,35605	0,05651	-0,64236	10,99860
$\hat{Y}_3$	-0,01667	0,00100	-0,07865	0,14114	0,15435	2,15291	6,27396
$\hat{Y}_4$	-0,00148	0,02853	-0,07973	-0,24374	0,05496	-0,67775	-0,71069
$\hat{Y}_5$	-0,00273	-0,01321	0,00909	-1,10723	0,16239	-1,01444	3,05804
$\hat{Y}_6$	0,00668	-0,01180	-0,10653	0,62460	0,00568	-1,01976	-3,34168

A próxima etapa foi a obtenção dos centróides para cada uma das classes temáticas. Estes centróides são mostrados na Tabela 18.

TABELA 18 - CENTRÓIDES DAS CLASSES TEMÁTICA PARA A GRADE 1m - TODAS AS VARIÁVEIS

Classe	$\hat{Y}_1$	$\hat{Y}_2$	$\hat{Y}_3$	$\hat{Y}_4$	$\hat{Y}_5$	$\hat{Y}_6$
AP	-4,19429	-0,67601	0,06531	1,05945	-1,10745	-0,45650
AV	9,71376	-0,43244	-2,92702	-0,14425	0,16760	-0,08922
BR	-6,44307	-1,47675	-0,73900	-2,24697	-1,47703	-0,94185
CS	-9,73763	-1,61292	-0,89741	1,05862	0,31061	0,09837
CI	-8,05585	-2,53219	-0,27211	-1,82339	0,74683	0,08154
ED	-8,64183	-1,75170	0,19608	0,67807	1,84186	0,74443
IN	2,29783	-1,54617	-1,96030	0,94960	-0,77288	0,65035
ME	-7,22889	0,80974	2,10092	-0,77939	-1,12157	0,49113
RE	6,46531	1,14688	1,05812	-0,17987	-0,24010	0,25328
AG	3,11055	4,99582	-1,06769	-0,49213	0,83286	-0,32465
SO	6,86751	-1,70786	3,03873	0,67772	0,93322	-0,44509
VA	9,42687	-2,11290	0,48898	-0,44114	-0,11842	-0,06097

A classificação da grade 1 metro ocorreu de forma similar a dos experimentos anteriormente realizados. A partir da classificação foi obtida a matriz de confusão que é mostrada na Tabela 19.

TABELA 19 - MATRIZ DE CONFUSÃO PARA A GRADE 1 m - TODAS AS VARIÁVEIS

Classe	AP	BR	CS	CI	ED	AG	VA	AV	ME	IN	RE	SO	Soma
AP	0,455	0,000	0,140	0,010	0,140	0,000	0,000	0,010	0,023	0,009	0,010	0,203	1,000
BR	0,000	0,623	0,002	0,000	0,000	0,000	0,000	0,032	0,021	0,039	0,282	0,001	1,000
CS	0,032	0,002	0,569	0,032	0,035	0,001	0,003	0,028	0,042	0,224	0,004	0,028	1,000
CI	0,012	0,023	0,007	0,394	0,029	0,008	0,014	0,057	0,034	0,405	0,007	0,009	1,000
ED	0,304	0,000	0,005	0,002	0,324	0,056	0,042	0,001	0,001	0,001	0,000	0,264	1,000
AG	0,042	0,000	0,001	0,008	0,038	0,384	0,272	0,004	0,001	0,002	0,000	0,248	1,000
VA	0,001	0,000	0,002	0,022	0,015	0,507	0,406	0,007	0,003	0,013	0,000	0,024	1,000
AV	0,023	0,035	0,008	0,035	0,014	0,002	0,002	0,483	0,315	0,081	0,001	0,001	1,000
ME	0,019	0,007	0,023	0,015	0,008	0,015	0,031	0,418	0,446	0,014	0,003	0,001	1,000
IN	0,019	0,012	0,208	0,185	0,012	0,005	0,003	0,121	0,108	0,273	0,050	0,004	1,000
RE	0,002	0,278	0,004	0,001	0,001	0,006	0,001	0,003	0,004	0,050	0,645	0,004	1,000
SO	0,172	0,013	0,001	0,014	0,056	0,013	0,034	0,002	0,015	0,022	0,002	0,656	1,000

Foram agrupados os índices de acerto para as classes obtidos no terceiro, quarto e quinto experimentos. A Tabela 20 apresenta estes resultados.

TABELA 20 - ÍNDICES DE ACERTO PARA AS CLASSES NO 3º, 4º E 5º EXPERIMENTO

CLASSE	3ºEXPERIMENTO	4ºEXPERIMENTO	5ºEXPERIMENTO
AP	0,834	0,659	0,455
ED	0,306	0,395	0,324
AG	0,625	0,524	0,384
SO	0,867	0,725	0,656
BR	0,855	0,732	0,623
CS	0,701	0,635	0,569
CI	0,703	0,576	0,394
VA	0,638	0,588	0,406
AV	0,768	0,654	0,483
ME	0,792	0,704	0,446
IN	0,754	0,498	0,273
RE	0,902	0,885	0,645

A partir da Tabela 20 foi obtido o gráfico mostrado na Figura 43.

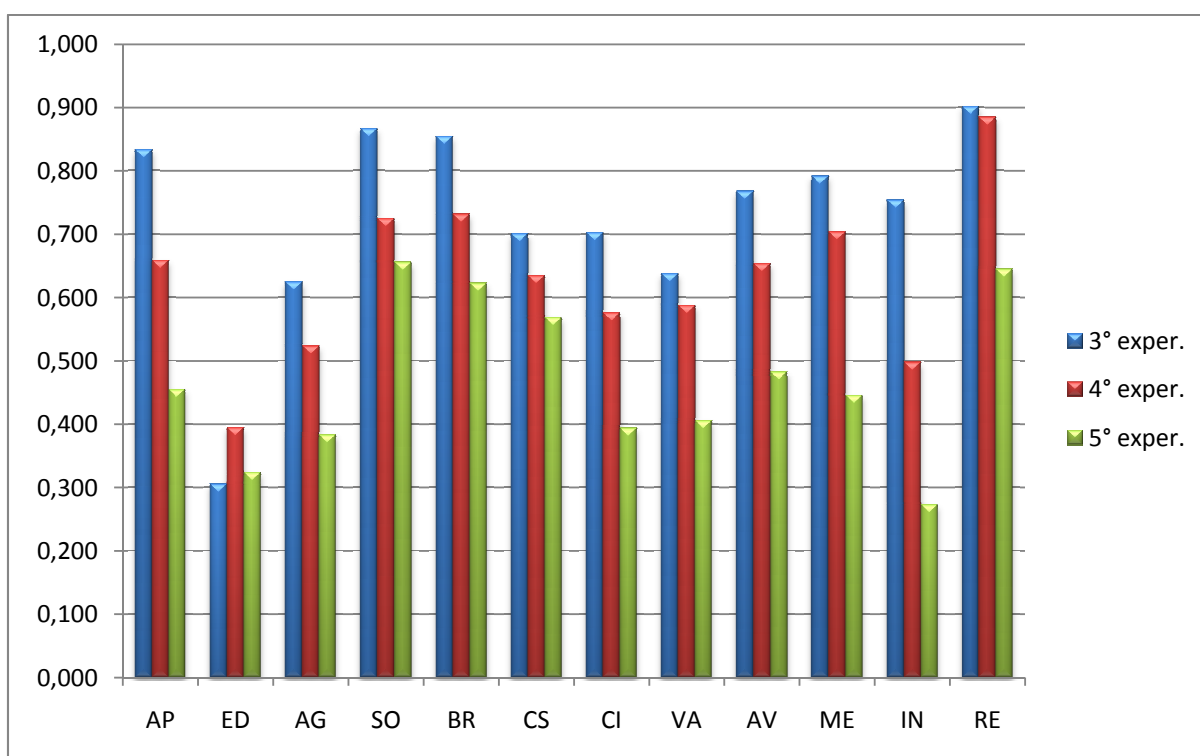


FIGURA 43 - ÍNDICES DE ACERTO PARA O 3º, 4º E 5º EXPERIMENTO

As seguintes considerações podem ser formuladas a partir dos resultados apresentados:

- Agropecuária (AP): o índice de acerto de 65,9% diminuiu para 45,5%. As maiores confusões foram verificadas com capoeirinha (CS) com 14,0%, edificações (ED) com 14,0% e solo exposto (SO) com 20,3%. As similaridades com estas classes foi aumentada com a adoção da grade de 1 metro e, portanto, não favorecem a separação da classe agropecuária (AP).

- Bracatinga (BR): neste experimento o índice de acerto sofreu um decréscimo de 11,1%, causado pelo aumento na confusão com a classe reflorestamento (RE) que passou de 24,7% para 28,2%. Também surgiram erros de classificação com as áreas ocupadas por floresta (AV), capoeirão (ME) e capoeira (IN). O índice de acerto na classificação para a classe bracatinga (BR) baixou de 73,2% para 62,3%.

- Capoeirinha (CS) : o índice obtido para a classificação desta classe foi de 56,9% inferior ao obtido no quinto experimento (63,5%). A confusão com a capoeira (IN) passou de 16,5% para 22,4%.

- Floresta de galeria (CI): o acerto para esta formação vegetal foi de 39,6% um valor 18% menor que o obtido no experimento anterior (57,6%). O principal motivo para esta piora está relacionado com o aumento da confusão com a classe capoeira (IN) que passou de 29,7% para 40,5%.

- Floresta (AV): houve redução no acerto do classificador para esta classe passando de 65,4% para 48,3%. O maior detalhamento obtido com a grade de 1 metro, fez com que a qualidade do MDT se tornasse ainda mais evidente. Prova disto é a confusão com a capoeirão (ME) passando de 16,2% no quarto experimento para 31,5% neste teste.

- Capoeirão (ME): conforme já constatado com a classe floresta (AV) a confusão aumentou. O valor no quarto teste foi de 14,3% e neste experimento aumentou para 41,8%. O índice de acerto diminuiu de 70,4% para 44,6%.

- Capoeira (IN): nesta classe o índice de acerto foi penalizado em 22,5%. O valor obtido no quarto experimento diminuiu de 49,8% para 27,3%. A confusão com a classe capoeirinha (CS) foi a que mais aumentou passando de 12,2% para 20,8%. A separação com as outras sucessões de formação vegetal também foi penalizada como pode ser percebido no gráfico apresentado na Figura 42.

- Reflorestamento (RE): o melhor desempenho nas formações vegetais foi conseguido com a classe reflorestamento (RE) com um índice de acerto igual a 64,5%. Comparando-se com o quarto experimento ocorreu um decréscimo de 24%.

A maior confusão ainda ocorre com a classe bracinga (BR) com um valor de 27,8%.

- Edificações (ED): a expectativa para esta classe seria um aumento do índice de acerto pois a pretensão seria resolver o problema das bordas. Contudo, isto acabou não ocorrendo. O valor de 39,5% obtido no quarto experimento baixou para 32,4%. A causa para este decréscimo é que a densidade de pontos LIDAR na área dos experimentos é de 2 pontos/m<sup>2</sup>. Como a resolução geométrica da grade regular é de 1 metro, a confusão com as classes agropecuária (AP) e solo exposto (SO) aumentou para, 30,4% e 26,4% respectivamente.

- Várzea (VA): o acerto na classificação da várzea (VA) foi penalizado, sendo obtido 40,6% de acerto contra 58,8% obtidos no experimento anterior. A confusão com a classe temática corpos d'água (AG) aumentou de 36,6% para 50,7% no quinto experimento.

- Corpos d'água (AG): o classificador obteve acerto de 38,4% valor inferior aos 52,4% verificados no quarto experimento. A maior confusão ainda é com a classe várzea (VA) com valor de 47,2%, superior ao índice obtido no quarto experimento que foi de 36,6%. No caso desta classe, ao analisar a imagem de intensidade foi possível verificar grande quantidade de sedimentos (areia) e cuja resposta ficou evidente quando da utilização da grade com resolução geométrica de 1 metro. O sedimentos presentes na água fizeram com que a confusão com a classe solo exposto (SO) subisse de 3,2% para 24,8%.

- Solo exposto (SO): o desempenho obtido pelo classificador diminui de 72,5% para 65,6% traduzido principalmente pelo aumento na confusão com a classe agropecuária que aumentou em 3,6%, passando de 14,6% no quarto experimento para 17,2% neste teste.

Analisando o desempenho na forma dos dois grupos antrópico e formação vegetal pode-se afirmar que:

- Para as formações naturais o aumento da resolução geométrica da grade penalizou ainda mais a qualidade da classificação. O índice médio de acerto para as classes baixou de 65,9% para 48,0%. As principais causas para a ocorrência disto são a baixa densidade de pontos LASER na célula da grade com resolução de 1 metro que ocasionou maior variabilidade dos parâmetros no interior das áreas de cada classe temática. Cita-se ainda como complicador a qualidade do MDT que prejudicou a separação dos estágios de sucessão vegetal. A classe menos

penalizada foi o capoeirinha (CS) com decréscimo de 6,6% quando comparado ao quarto experimento. A formação vegetal mais penalizada foi capoeirão (ME) com decréscimo no índice de acertos de 25,8% reforçando a falta da qualidade do MDT, pois confusão com os demais estágios de sucessão vegetal aumentou.

- Para o grupo composto pelas classes edificação (ED), solo exposto (SO), agropecuária (AP) e corpos d'água (AG) o índice médio de acerto baixou de 57,6% para 45,5% no quinto experimento. A classe relativa a agropecuária (AP) sofreu a maior perda no valor de acertos com um decréscimo de 20,4%. A classe edificação (ED), ao contrário do que se esperava, obteve índice de acerto inferior ao verificado no quarto experimento baixando de 39,5% para 32,4%.

O quinto experimento apresentou acerto médio de 47,1%, valor inferior ao conseguindo no quarto teste que foi de 65,9%. Analisando o gráfico mostrado na Figura 43 verifica-se que o aumento da resolução geométrica da grade regular não acarretou melhorias ao processo de classificação e, portanto, as grades regulares de 5 metros e 1 metros não serão empregadas nos testes com as Funções Discriminante Quadrática e com a Regressão Logística com Camada Oculta (RLO). Destaca-se mais uma vez que os dois motivos que prejudicaram o desempenho da Função Discriminante Linear no quarto e quinto experimentos foram:

- A baixa densidade de pontos acabou ocasionando uma maior variabilidade dos parâmetros no interior das células da grade, que compõe as diferentes áreas para as classes temáticas analisadas. Isto se traduz em uma maior heterogeneidade nas células internas aos limites das formações vegetais ocasionando o aumento da confusão entre as classes temáticas.

- A qualidade do MDT ficou evidente, pois a separação entre as formações vegetais foi bastante penalizada. Deve ser lembrado que a altura média da cobertura vegetal foi calculada a partir da diferença entre a altitude do ponto e altitude do plano ajustado para cada célula de 10 metros. O tamanho da plano ajustado também refletiu sobre este quesito.

Para o sexto experimento foi realizada a classificação da grade regular de 10 metros empregando-se todos os cofatores e a Função Discriminante Quadrática.

#### 4.4.2 Classificação Empregando a Função Discriminante Quadrática

Visando comparar os resultados obtidos com a Função Discriminante Linear de Fisher, o sexto experimento empregou a Função Discriminante Quadrática com o emprego de todas as variáveis na classificação da grade de 10 metros.

O ajuste das funções de classificação ocorreu de maneira similar: entrada dos valores dos cofatores e cálculo dos coeficientes das 12 funções.

O primeiro passo foi realizar o ajuste das funções quadráticas. As funções quadráticas dependem de coeficientes de segundo grau, primeiro grau e das constantes. A metodologia adotada para o ajuste das funções foi a mesma utilizada por Paixão (2006).

Os resultados obtidos com o ajuste das funções quadrática é apresentado no apêndice deste trabalho.

A partir das funções de discriminação foi implementada a classificação da grade de 10 metros e então gerada a respectiva matriz de confusão. Esta matriz é mostrada na Tabela 21.

**TABELA 21 - MATRIZ DE CONFUSÃO PARA A GRADE 10 m - TODAS AS VARIÁVEIS - FUNÇÃO DISCRIMINANTE QUADRÁTICA**

Classe	AP	BR	CS	CI	ED	AG	VA	AV	ME	IN	RE	SO	Soma
AP	0,773	0,000	0,051	0,029	0,026	0,000	0,000	0,023	0,032	0,030	0,000	0,037	1,000
BR	0,000	0,824	0,000	0,000	0,000	0,000	0,000	0,006	0,001	0,003	0,166	0,000	1,000
CS	0,024	0,000	0,702	0,023	0,008	0,001	0,002	0,052	0,034	0,149	0,000	0,005	1,000
CI	0,001	0,000	0,005	0,616	0,000	0,000	0,000	0,058	0,044	0,267	0,009	0,000	1,000
ED	0,310	0,000	0,008	0,003	0,363	0,000	0,000	0,020	0,006	0,000	0,000	0,290	1,000
AG	0,000	0,000	0,007	0,006	0,030	0,578	0,342	0,009	0,021	0,000	0,000	0,007	1,000
VA	0,011	0,000	0,000	0,017	0,042	0,302	0,620	0,004	0,000	0,000	0,000	0,004	1,000
AV	0,004	0,000	0,038	0,016	0,000	0,000	0,000	0,757	0,079	0,095	0,010	0,000	1,000
ME	0,001	0,003	0,033	0,015	0,004	0,000	0,000	0,127	0,805	0,007	0,000	0,005	1,000
IN	0,015	0,004	0,064	0,159	0,010	0,000	0,000	0,062	0,074	0,609	0,001	0,002	1,000
RE	0,000	0,175	0,008	0,000	0,000	0,000	0,000	0,011	0,003	0,001	0,802	0,000	1,000
SO	0,181	0,000	0,000	0,005	0,002	0,003	0,000	0,000	0,002	0,000	0,001	0,806	1,000

A partir dos índices de acerto para as classes temáticas obtidos no terceiro e sexto experimentos, elaborou-se a Tabela 22.



TABELA 22 - ÍNDICES DE ACERTO PARA AS CLASSES NO 3º E 6º EXPERIMENTO

CLASSE	3º EXPERIMENTO	6º EXPERIMENTO	INCREMENTO
AP	0,834	0,773	-0,061
ED	0,306	0,253	-0,053
AG	0,625	0,578	-0,047
SO	0,867	0,747	-0,120
BR	0,855	0,824	-0,031
CS	0,701	0,732	0,031
CI	0,703	0,667	-0,036
VA	0,638	0,620	-0,018
AV	0,768	0,757	-0,011
ME	0,792	0,805	0,013
IN	0,754	0,609	-0,145
RE	0,867	0,802	-0,065

Os valores mostrados na Tabela 22 possibilitaram a obtenção do gráfico mostrado na Figura 44.

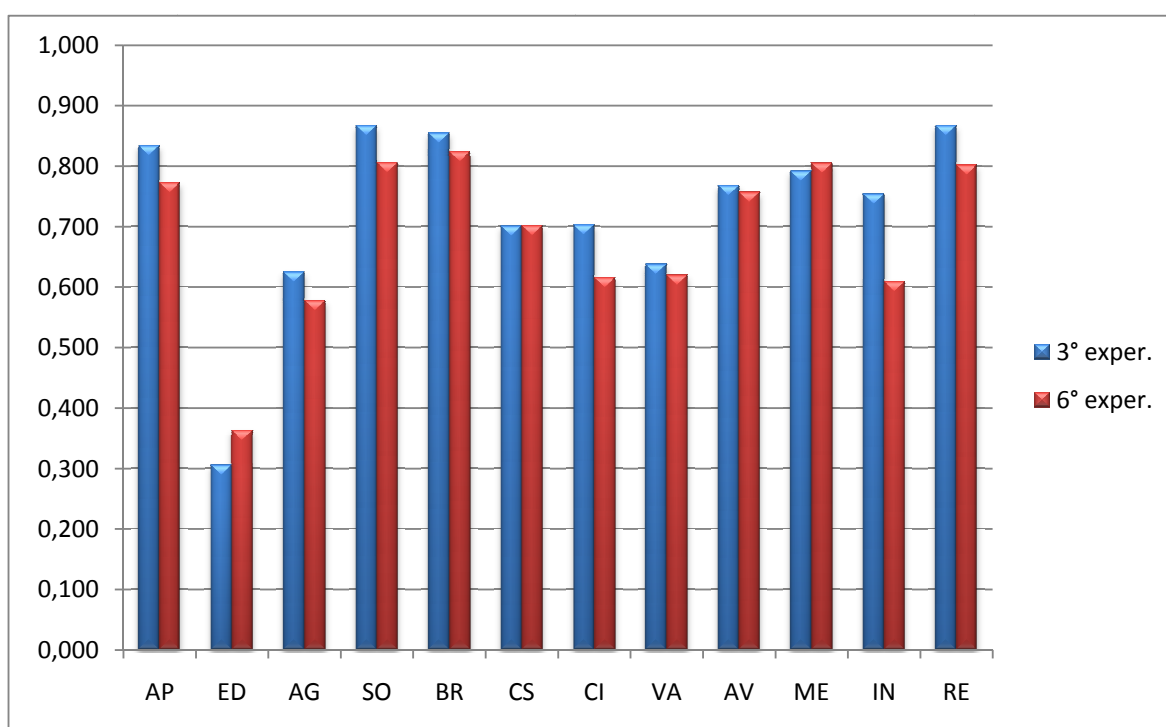


FIGURA 44 - ÍNDICES DE ACERTO PARA O 3º E 6º EXPERIMENTO

Antes do início da análise dos resultados, cabe lembrar que a diferença entre a FDL e a Função Discriminante Quadrática é que no primeiro caso as

matrizes de co-variâncias são consideradas iguais e no segundo consideram-se as matrizes de co-variâncias das classes diferentes. Com a confrontação dos resultados obtidos no 3º e 6º experimento pode-se afirmar que:

- Agropecuária (AP): o resultado da classificação desta classe para o terceiro experimento foi de 83,4%, valor superior ao alcançado quando se empregou a Função Discriminante Quadrática para a qual o índice de acerto foi de 77,3%. Este decréscimo no índice de acerto está relacionado com o aumento da confusão com as classes capoeirinha (CS) em 5,1% e com as formações naturais floresta (AV), capoeirão (ME) e capoeira (IN) para as quais a confusão foi de 2,3%, 3,2% e 3,0%. A causa para esta piora está na forma como a Função Quadrática delimita as classes no espaço canônico. A separação assumindo uma forma quadrática mostrou-se menos eficiente que a forma linear quando se tentou separar as diferentes sucessões das formações naturais.

A Função Discriminante Quadrática mostrou-se, no entanto, mais eficiente na distinção com as classes edificação (ED) e solo exposto (SO). No caso das edificações (ED) a Função Discriminante de Fisher obteve um índice de confusão de 5,3%, valor superior ao obtido pela Função Discriminante Quadrática que foi de 2,6%. Quando se observa a confusão com solo exposto (SO) também se verifica melhoria, com o erro diminuindo de 8,8% para 3,7%. Isso significa que a separação não linear disponibilizada pela Função Discriminante Quadrática foi mais eficiente para estas classes temáticas.

- Bracatinga (BR): o índice de acerto obtido no terceiro experimento para esta classe foi de 85,5%, índice 3,1% superior ao valor obtido no sexto teste (82,4%). A diferença de 3,1% entre os classificadores mostra o melhor desempenho da função de Fisher para a separação da classe bracatinga (BR). A confusão com a classe reflorestamento (RE) obtida neste experimento foi de 16,6%, valor superior ao obtido com a FDL (12,8%). O motivo para esta confusão está relacionado com a forma de separação das classes, na qual a separação linear mostrou ser a mais apropriada para a classe bracatinga (BR).

- Capoeirinha (CS): o acerto desta classe foi de 70,2%, praticamente igual ao obtido com a FDL (70,1%). A confusão com a classe capoeirinha permanece em 14,9%, valor também similar ao obtido no terceiro experimento (14,8%). A Função Discriminante Quadrática mostrou-se novamente mais eficiente na separação das edificações (ED) e solo exposto (SO) obtendo índices de confusão de 0,8% e 0,5%

respectivamente. A tendência de um resultado pior na separação das formações vegetais se confirmou pois o índice de confusão da capoeirinha (CS) com a floresta (AV) aumentou de 1,4% para 5,8%.

- Floresta de galeria (CI): esta classe foi classificada corretamente em 66,7%, índice de acerto inferior ao obtido no terceiro experimento que foi de 70,3%. A confusão com as áreas ocupadas por florestal (AV) aumentou de 2,8% para 5,8% e a confusão com a capoeira passou de 24,7% para 26,7%.

- Floresta (AV): o índice de acerto para esta classe foi de 75,7%, valor inferior ao obtido no terceiro experimento que foi de 76,8%. A maior confusão é percebida com a capoeira (IN) com um valor de 9,5%, superior aos 6,8% obtidos no terceiro experimento. A confusão com as demais formações naturais ainda pode ser verificada.

- Capoeirão (ME): o índice de acerto nesta formação foi de 80,5% superior ao obtido no terceiro teste com valor de 79,2%. A maior confusão ainda ocorre com a floresta (AV) com um índice de 12,7%, superior aos 9,3% obtidos quando empregada a FDL.

- Capoeira (IN): para esta formação natural o classificador obteve um acerto de 60,9%, um decréscimo de 14,5% quando comparado com o resultado do terceiro experimento (75,4%). O maior erro de classificação ocorreu com a classe capoeira com 15,9%, valor superior ao conseguido no terceiro experimento (7,5%).

- Reflorestamento (RE): o índice de acerto para esta classe foi de 80,2% valor 10,0% menor que o obtido com a FDL. O maior índice de confusão ocorreu com a classe bracatinga (BR) verificando-se o índice de 17,5%, sendo esta a principal causa para o decréscimo no acerto desta classe.

- Edificações (ED): o acerto do classificador para as edificações foi de 36,3% valor 5,7% maior que o do terceiro experimento (30,6%). A maior confusão ainda persiste com a classe agropecuária (AP) com um índice de 31,0% e com a classe solo exposto (SO) com 29,0%. Contudo, estes índices de confusão são respectivamente 4,3% e 3,4% menores que os obtidos quando se empregou a FDL.

- Várzea (VA): foi classificada com acerto de 62,0% valor 1,8% menor que o do terceiro experimento. A maior confusão é percebida com os corpos d'água (AG) com um valor de 30,2%.

- Corpos d'água (AG): o índice de acerto para esta classe foi de 57,8% valor inferior quando comparado ao terceiro experimento (62,5%). A confusão com a classe várzea (VA) foi de 34,2%.

- Solo exposto (SO): o desempenho da classificação do solo exposto foi de 80,6%, valor 6,1% menor que o índice de 86,7% obtido quando empregada a FDL. A maior confusão ocorre ainda com a classe agropecuária (AP) com um índice de confusão 18,1%. Porém, obteve-se uma melhor separação das edificações (ED) com uma confusão de 0,2%.

Analisando os resultados para o grupo de classes relativas a cobertura vegetal pode-se afirmar que:

- O melhor desempenho foi alcançado através da classificação da bracatinga (BR) com um acerto de 82,4%.

- O menor desempenho foi obtido para a classe capoeira (IN) obtendo-se um índice de acerto igual a 60,2%.

- Quando comparadas com os resultados da classificação obtida com a Função Discriminante de Fisher ou Linear percebe-se que os resultados da Função Discriminante Quadrática foram inferiores em até 14,5%. Este menor desempenho com relação às formações naturais está diretamente ligado a separação não linear entre classes propiciada pela Função Discriminante Quadrática. É possível afirmar que o classificador linear é mais eficiente na separação das formações vegetais. A prova deste fato pode ser constatada pelo acerto médio para a cobertura vegetal que foi de 76,4%, no terceiro experimento, contra 71,7% no sexto experimento.

- Para as quatro classes: solo exposto (SO), agropecuária (AP), corpos d'água (AG) e edificações (ED) o acerto médio obtido foi de 63,0% valor ligeiramente inferior aos 65,8% obtidos com o terceiro teste. Apesar do menor índice de acerto percebe-se que a separação não linear de classes gerou menos confusão entre as classes deste grupo ao comparar com os erros obtidos na classificação empregando FDL. Destaca-se ainda o desempenho superior na classificação das edificações com melhoria de 5,7% no índice de acerto, obtendo-se um valor de 36,3%.

O acerto médio do sexto experimento foi de 68,8%, 4,1% menor que os 72,9% obtidos com o terceiro experimento. Pode-se afirmar que as mesmas causas de confusão detectadas no terceiro experimento afetaram o desempenho da classificação com a Função Discriminante Quadrática. Percebe-se, no entanto, que o modelo de separação de classes não linear é mais apropriado para as classes do

grupo antrópico percebendo-se o oposto para as formações naturais. Outro aspecto negativo que deve ser considerado é o esforço computacional necessário para obter a classificação empregando o discriminante quadrático.

#### 4.4.3 Classificação Empregando a RLO

Para o sétimo experimento foi procedida a classificação da grade regular de 10 metros empregando-se a Regressão Logística com Camada Oculta (RLO). Empregaram-se os seis cofatores: altitude média, intensidade média, altura média da cobertura vegetal, desvio padrão da altitude, desvio padrão da intensidade e densidade de pontos.

Como grupo de referência empregou-se a primeira classe armazenada no arquivo de amostras (agropecuária). Em termos operacionais, esta classificação exigiu mais esforço computacional para a sua implementação quando comparada à Função Discriminante de Fisher. Isto se deve principalmente ao método iterativo necessário ao cálculos dos estimadores de verossimilhança ( $\beta$ ).

A partir das amostras foram obtidas as derivadas parciais mostradas na equação (79). Os estimadores ( $\beta$ ) das derivadas são calculados iterativamente. Os estimadores são recalculados empregando-se os valores de  $\hat{Y}$  (camada oculta). A classificação da célula ocorre comparando a probabilidade obtida com a probabilidade do grupo de referência, isto é, o cálculo das probabilidades condicionadas. A menor das dozes probabilidades indica a qual classe pertence a célula analisada.

A partir das amostras foram calculados os coeficientes das equações de verossimilhança apresentados na Tabela 23.

TABELA 23 - COEFICIENTES DAS FUNÇÕES DE DISCRIMINAÇÃO PARA A GRADE 10 m - TODAS AS VARIÁVEIS - RLO

Classe	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$	$\beta_6$
1	-64,7047	0,0737	0,0676	-0,5628	0,6681	-0,0499	0,9702
2	-151,1647	0,2017	-0,0579	-0,4774	-0,7335	-0,197	-1,9864
3	-117,8472	0,1707	-0,0521	-0,8538	1,0613	-0,1973	-6,7161
4	-41,37	0,067	-0,1187	-0,521	0,3272	-0,0474	1,2947
5	-68,4879	0,1083	-0,1587	-0,2374	-0,0364	-0,4088	2,8676
6	-54,464	0,092	-0,1474	-0,8771	-2,0679	-0,1092	0,837
7	-105,1118	0,1561	-0,1924	-1,8845	-0,2725	0,1477	-2,5841
8	-131,1016	0,1571	-0,0209	-0,5859	1,8658	-0,0315	2,7992
9	-72,6465	0,0953	-0,0481	-0,1912	0,5047	0,1138	-2,9707
10	-113,4275	0,1325	-0,067	-1,1201	-0,1856	0,2281	5,1791
11	-150,6264	0,211	-0,2341	-1,2952	0,3189	-0,0618	1,7485

A partir das equações de verossimilhança procedeu-se a classificação das células da grade de 10 metros. De maneira similar aos experimentos realizados anteriormente, obteve-se a matriz de confusão que é mostrada na Tabela 24.

TABELA 24 - MATRIZ DE CONFUSÃO PARA A GRADE 10 m - TODAS AS VARIÁVEIS - RLO

Classe	AP	BR	CS	CI	ED	AG	VA	AV	ME	IN	RE	SO	Soma
AP	0,856	0,000	0,007	0,000	0,008	0,000	0,000	0,000	0,001	0,004	0,000	0,124	1,000
BR	0,000	0,895	0,000	0,000	0,000	0,000	0,002	0,005	0,000	0,003	0,095	0,000	1,000
CS	0,046	0,005	0,784	0,071	0,003	0,000	0,000	0,020	0,010	0,035	0,000	0,026	1,000
CI	0,000	0,000	0,008	0,845	0,000	0,001	0,001	0,013	0,008	0,119	0,000	0,006	1,000
ED	0,231	0,000	0,002	0,000	0,579	0,009	0,008	0,000	0,006	0,001	0,002	0,262	1,000
AG	0,004	0,000	0,003	0,000	0,009	0,560	0,410	0,005	0,001	0,005	0,000	0,004	1,000
VA	0,000	0,000	0,001	0,003	0,000	0,377	0,582	0,000	0,000	0,019	0,000	0,019	1,000
AV	0,000	0,014	0,060	0,032	0,002	0,000	0,005	0,764	0,092	0,031	0,000	0,000	1,000
ME	0,003	0,000	0,004	0,008	0,000	0,000	0,000	0,141	0,761	0,082	0,000	0,001	1,000
IN	0,006	0,008	0,018	0,009	0,000	0,007	0,010	0,135	0,014	0,766	0,009	0,018	1,000
RE	0,000	0,073	0,008	0,000	0,000	0,000	0,000	0,001	0,000	0,002	0,916	0,000	1,000
SO	0,079	0,000	0,003	0,005	0,006	0,003	0,000	0,000	0,000	0,006	0,000	0,898	1,000

Para realizar a comparação do sétimo com o terceiro experimento elaborou-se a Tabela 25, nas quais podem ser apreciados os índices de acerto para as classes temáticas.

TABELA 25 - ÍNDICES DE ACERTO PARA AS CLASSES NO 3º E 7º EXPERIMENTO

CLASSE	3ºEXPERIMENTO	7ºEXPERIMENTO	INCREMENTO
AP	0,834	0,856	0,022
ED	0,306	0,479	0,173
AG	0,625	0,560	-0,065
SO	0,867	0,898	0,031
BR	0,855	0,895	0,040
CS	0,701	0,784	0,083
CI	0,703	0,845	0,142
VA	0,638	0,582	-0,056
AV	0,768	0,764	-0,004
ME	0,792	0,761	-0,031
IN	0,754	0,766	0,012
RE	0,867	0,916	0,049

O gráfico mostrado na Tabela 45 apresenta o desempenho da classificação da grade regular de 10 metros quando se empregaram os três classificadores propostos e a configuração de variáveis mencionada.

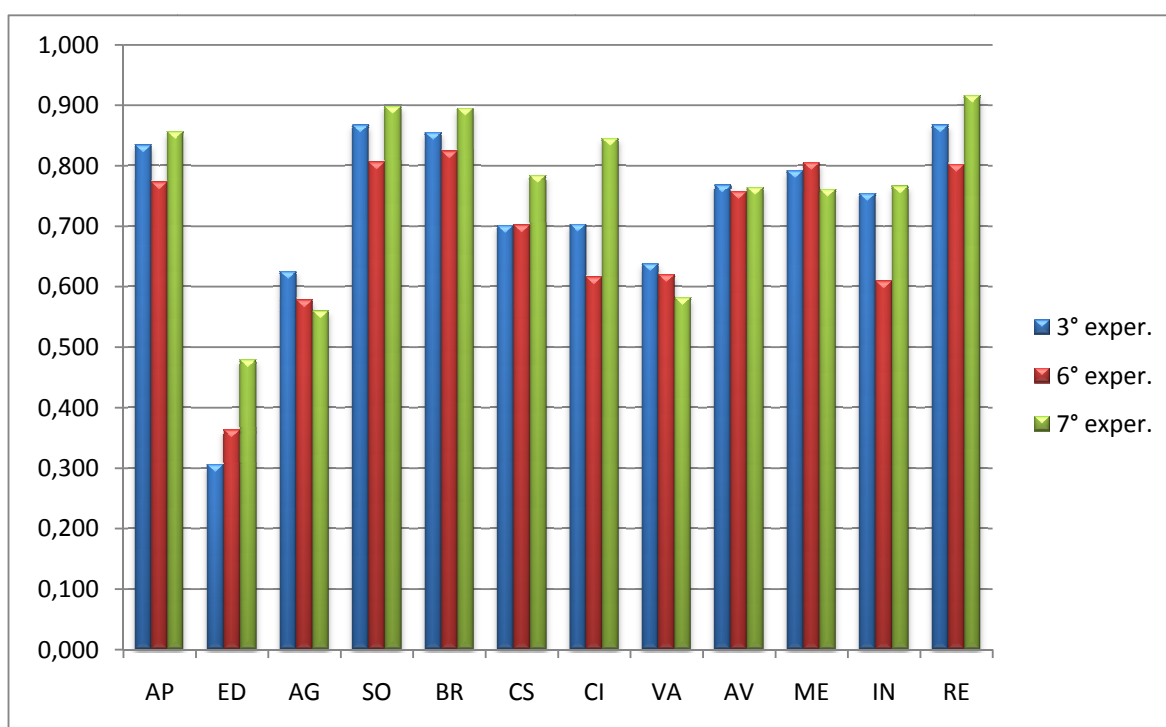


FIGURA 45 - ÍNDICES DE ACERTO PARA O 3º, 6º E 7º EXPERIMENTO

A análise preliminar dos resultados mostrados na Tabela 24 e da Figura 45 possibilitam afirmar que os mesmos problemas relacionados a confusão das classes

e que foram verificados quando da realização do terceiro experimento também afetaram a classificação da grade regular de 10 metros com o emprego da RLO. Isto posto, pode-se propor as seguintes considerações:

- Agropecuária (AP): o desempenho da RLO na separação da classe agropecuária foi de 85,6%. Este índice é superior em 2,2% ao índice obtido com o emprego da Função Discriminante de Fisher. A confusão com a classe capoeirinha (CS) diminuiu de 2,3% para 0,7%. A maior confusão ainda ocorre com a classe solo exposto (SO) com um índice de 12,4%, que é 3,6% maior que o obtido no terceiro experimento cujas causas foram mencionadas anteriormente.

- Bracatinga (BR): o índice de acerto obtido no sétimo experimento para esta classe foi de 89,5%, índice 4,0% superior ao valor obtido no terceiro teste (85,5%). A RLO apresenta vantagem na separação da classe bracatinga (BR) o que pode ser reforçado pela diminuição do valor da confusão com a classe reflorestamento (RE) que passou de 12,8% para 9,5%.

- Capoeirinha (CS) : o incremento no percentual de acertos da classe capoeirinha (CS) foi de 8,3% fazendo com que o índice de acerto do terceiro experimento subisse de 70,1% para 78,4% no sétimo experimento.

- Floresta de galeria (CI): a classificação desta formação natural obteve um acerto de 84,5% superior aos 70,3% verificados com a FDL. A melhoria deste índice está ligada diretamente ao melhor delineamento da capoeira (IN) cujo índice de confusão baixou de 24,7% para 11,9%.

- Floresta (AV): a classificação desta formação natural obteve um índice de acerto inferior ao verificado no terceiro experimento. Este índice passou de 76,8% para 76,4% um decréscimo de 0,4%. Cabe ressaltar que a confusão com as sucessões vegetais está relacionada com a qualidade do MDT, conforme já mencionado.

- Capoeirão (ME): o índice de acerto nesta formação foi de 76,1% valor inferior aos 79,2% quando realizada a classificação com a Função Discriminante de Fisher. A maior confusão é percebida com a classe floresta (AV) com um índice de 14,7%, superior aos 9,3% obtidos quando empregada a FDL. Novamente, a qualidade do MDT prejudicou a classificação desta formação natural.

- Capoeira (IN): a classificação desta classe apresentou melhoria de 1,2% em relação ao terceiro experimento. A maior confusão pode ser percebida com a classe temática floresta (AV).



- Reflorestamento (RE): o índice de acerto para esta classe foi de 91,6% valor 0,6% maior que o obtido com a FDL. O índice de confusão com a bracatinga (BR) ainda persiste com um valor de 7,3%.

- Edificações (ED): o maior incremento no índice de acerto foi obtido para a classe edificações. Neste caso, a RLO conseguiu classificar corretamente 47,9% das edificações, ou seja, um incremento de 17,3% na taxa de acertos. A diminuição de 2,8% na confusão com a classe agropecuária (AP) e de 11,2% com a classe solo exposto (SO) auxiliaram na melhoria da taxa de acertos para a edificação.

- Várzea (VA): foi corretamente classificada em 58,2% dos casos. A maior confusão ainda ocorre com a classe corpos d'água (AG) com um valor de 37,7%. Neste caso, a RLO apresentou desempenho inferior aos 63,8% alcançados pela Função Discriminante de Fisher.

- Corpos d'água (AG): o índice de acerto para esta classe foi de 56,0% valor inferior aos 62,5% obtidos com o uso da Função Discriminante de Fisher. A confusão com a classe várzea (VA) foi de 41,0%, ou seja, 10% maior que a verificada no terceiro experimento.

- Solo exposto (SO): o desempenho na classificação do solo exposto foi de 89,8%, valor 3,1% maior que o obtido com o uso da FDL. A maior confusão ocorreu com a classe agropecuária (AP) com um valor de 7,9%.

Quando consideradas as classes: bracatinga (BR), capoeirinha (CS), floresta de galeria (CI), várzea (VA), floresta (AV), capoeirão (ME), capoeira (IN) e reflorestamento (RE) pode-se concluir que:

- O melhor desempenho foi alcançado com a classificação das áreas de reflorestamento (RE) com uma taxa de acerto de 91,6%.

- O menor desempenho para o grupo das formações naturais foi percebido na classe várzea (VA) com 58,6% de acerto.

- Comparando os resultados da classificação obtidos com a Função Discriminante de Fisher e a RLO obteve-se uma melhora de até 14,2%, ocorrido para a floresta de galeria (CI). As formações naturais várzea (VA), floresta (AV) e capoeirão (ME) apresentaram decréscimo respectivamente de 5,6%, 0,4% e 3,1%. O acerto médio para o grupo de classes formação natural foi de 78,9%, superior aos 76,4% obtidos no terceiro experimento.

Para as quatro classes: solo exposto (SO), agropecuária (AP), corpos d'água (AG) e edificações (ED) o acerto médio obtido foi de 69,8% valor 4,0% superior ao índice médio verificado no terceiro teste.

Os resultados obtidos revelam que a Regressão Logística com Camada Oculta apresenta uma pequena vantagem na separação das classes quando se compara com o desempenho da Função Discriminante Linear de Fisher. Verifica-se que o acerto médio do sétimo experimento foi de 75,9%, portanto, é 3,0% maior que índice médio de 72,9% obtido no terceiro experimento.

Ao confrontar os resultados obtidos no sétimo experimento com a Função Discriminante Quadrática verifica-se que:

- O acerto médio na separação das classes relativas às formações naturais é 7,2% maior com o emprego da Regressão Logística com Camada Oculta. A maior diferença é verificada na separação da classe temática Floresta de Galeria (CI) com índice de acerto 17,8% superior no sétimo experimento. Contudo, na separação da classe capoeira (ME), a Função Discriminante Quadrática obteve um índice de acerto 4,4% superior ao índice obtido no sétimo experimento;
- Os resultados da separação das classes antrópicas (agropecuária (AP), solo exposto (SO), corpos d'água (AG) e edificações (ED)) mostra vantagem para a Regressão Logística com Camada Oculta, com um índice de acerto superior em 6,8% quando comparado com o resultado do sexto experimento;
- O índice de acerto do sétimo experimento foi 7,1% superior ao índice obtido quando se empregou a Função Discriminante Quadrática.

Com relação ao tempo de processamento computacional necessário ao ajuste das funções e classificação da grade LASER, foram verificados:

- Terceiro experimento: 14,617 minutos;
- Sexto experimento: 27,850 minutos e
- Sétimo experimento: 32,117 minutos.

O tempo de processamento computacional da Função Discriminante Linear foi 2,19 vezes menor que o necessário para a Regressão Logística com Camada Oculta. Comparando-se a Função Discriminante Linear e a Quadrática, a primeira, é 1,90 mais rápida que a segunda função. O motivo para a diferença dos tempos de processamento é que os métodos do sexto e sétimo experimentos são iterativos e para a Função Discriminante Linear o ajuste é obtido sem iterações.

A diferença nos tempos de processamento computacional pode-se tornar um fator relevante na escolha da função de classificação, à medida que a área de classificação aumente.

Ainda deve ser mencionado que a Regressão Logística com Camada Oculta é uma alternativa para a discriminação de classes empregando a Função Discriminante de Fisher, especialmente quando a condição de matrizes de covariâncias iguais não é satisfeita, além de independender da distribuição de probabilidade dos dados.

A matriz de classificação obtida no 7º experimento foi então utilizada para a criação de um mapa temático. Para obtenção do mapa mencionado foram empregadas as ferramentas disponíveis para tratamento de informações matriciais do programa Arc Gis. A imagem temática criada foi então convertida para o formato vetorial, gerando-se polígonos para as classes de cobertura do solo.

A variável visual escolhida para o mapa temático qualitativo foi o tom de cor. O mapa obtido neste processo pode ser visualizado Figura 46.

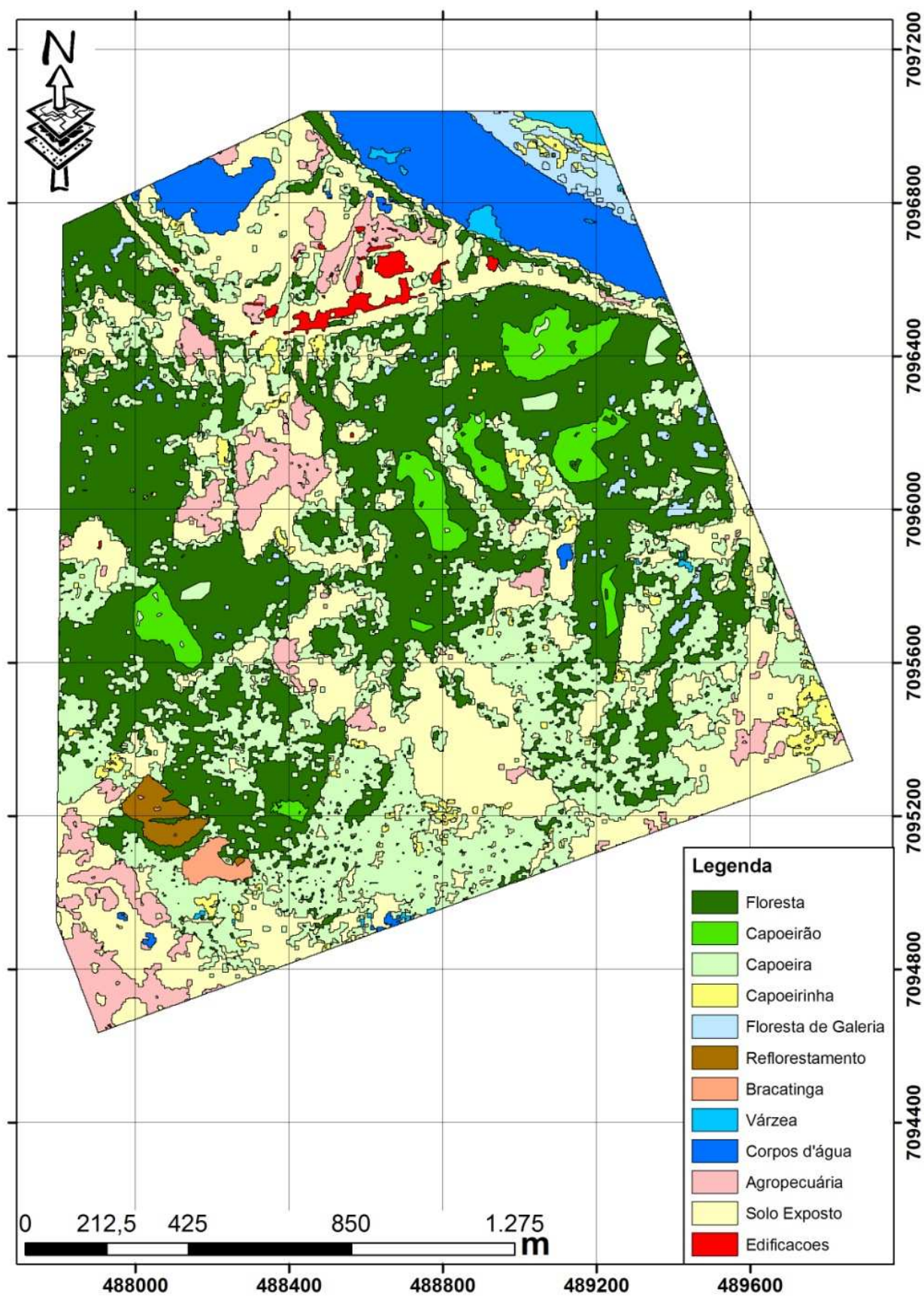


FIGURA 46 - CLASSIFICAÇÃO DA COBERTURA DO SOLO OBTIDA COM A RLO - GRADE 10m

## 5 CONCLUSÕES E RECOMENDAÇÕES

O presente trabalho de pesquisa teve como principal objetivo o desenvolvimento de metodologias que visam à classificação da cobertura do solo a partir de dados de um levantamento *laser scanner* e função discriminante provenientes da análise estatística multivariada.

Para a obtenção das informações necessárias para a classificação da cobertura do solo fez-se necessário o desenvolvimento de algoritmo para a filtragem dos dados LASER visando a obtenção de pontos que permitissem a modelagem matemática da superfície do terreno (MDT).

Com base nos resultados dos experimentos realizados são apresentadas as seguintes conclusões:

- Os experimentos foram realizados com êxito, comprovando que a metodologia é capaz de obter a classificação dos dados LASER de acordo com a cobertura do solo detectada pela restituição fotogramétrica e reambulação de campo. A automação obtida para a classificação da grade LASER é um dos resultados positivos alcançados nesta pesquisa.

- Para a obtenção dos dados necessários à geração do MDT, foi necessário o desenvolvimento computacional de um filtro morfológico adaptado. Este filtro conseguiu separar os pontos em duas classes: pontos sobre o terreno e sobre as elevações. Visando verificar os resultados obtidos no processo de filtragem matemática analisaram-se duas áreas menores com cobertura do solo distintas, além de toda a área onde foi testada o filtro. A comparação, entre as superfícies interpoladas à partir dos dados obtidos pelo processo de filtragem e fotogrametria, para a área um (ocupada por formações naturais e agropecuária) revelou a ocorrência de um rebaixamento médio de 1,294 m. Na área dois (ocupada por formações naturais) observou-se um rebaixamento de 2,653 m. A comparação das superfícies, ao ser estendida para toda a área de estudo, revelou que a tendência de rebaixamento permaneceu, resultando em um valor médio de 1,518 m. Neste contexto, deve ser destacado que mesmo sendo obtido por um fotogrametrista experiente, os dados da fotogrametria evidenciaram, especialmente em áreas cobertas por florestas, o problema de contato da marca flutuante com o terreno, que

também é estimado pelo fotogrametrista nestas áreas. Apesar de constatar-se a necessidade de um aprimoramento do filtro morfológico, este atingiu as expectativas iniciais, conseguindo separar os pontos, mesmo em áreas cobertas por florestas, nas classes almejadas. Destaca-se que a definição do tamanho da célula para ajuste dos planos locais foi a atividade que mais consumiu tempo nesta etapa da pesquisa. Os resultados alcançados na filtragem matemática motivam o aprimoramento do filtro proposto, pois as soluções atualmente disponíveis se restringem a aplicações comerciais. A biblioteca de aplicativos desenvolvidos para a filtragem se constitui em ponto de partida para o aprimoramento desta ferramenta.

A análise dos resultados alcançados com a separação das classes relativas à cobertura do solo permite concluir que:

- Os experimentos para a obtenção das classes temáticas foram realizados com êxito, comprovando que a metodologia proposta é capaz de obter a classificação dos dados LASER de acordo com a cobertura do solo detectada pela reambulação de campo e restituição fotogramétrica.

- O *laser scanner* é uma tecnologia emergente e seu emprego na obtenção de informações altimétricas está bastante consolidado. Carecem ainda, metodologias que propiciem o melhor aproveitamento da tecnologia, justamente na classificação da cobertura do solo das áreas imageadas. O desenvolvimento das metodologias de classificação apresentadas, ajudam a suprir esta lacuna conseguindo sucesso significativo na separação das classes temáticas relativas a cobertura do solo, destacando-se que foram utilizados somente os dados LASER no processo. O sucesso da metodologia é perceptível, pois obteve-se um índice de acerto médio na separação das classes temáticas referentes a cobertura do solo de até 75,9%.

- A escolha das variáveis (cofatores) constituiu-se em atividade primordial para as atividades de classificação das formações vegetais a partir das células da grade regular. A utilização de todos os cofatores analisados: altitude média, desvio padrão da altitude, intensidade média, desvio padrão da intensidade, altura média da cobertura vegetal e densidade de pontos, é a melhor configuração para a separação das classes temáticas. O cofator altura média da cobertura vegetal mostrou-se um

importante parâmetro na separação das classes temáticas relativas as sucessões das formações vegetais.

- A resolução geométrica de 10 metros para a grade regular foi a mais adequada para a classificação da cobertura vegetal na área de estudos. A utilização de grades regulares com maior resolução geométrica acabou por prejudicar as taxas de acerto da Função Discriminante de Fisher. Isto se deve ao fato de que ocorreu uma maior variabilidade no interior das áreas das classes temáticas e que não foi adequadamente modelada pelos cofatores selecionados. A menor densidade de pontos foi um fator decisivo para esta variabilidade, o que ficou evidenciado nos resultados obtidos com a classificação da grade regular com resolução geométrica de 1 metro, quando verificou-se que o índice de acerto médio resultou em 47,1%.

- O sucesso na separação das sucessões de regeneração da vegetação natural não foi mais promissor devido à qualidade do MDT obtido pelo processo de filtragem. A configuração da varredura LASER, de forma a aumentar a densidade dos pontos, bem como, o aprimoramento do processo de filtragem matemática serão fatores fundamentais para aumentar o índice de acerto na separação das sucessões das formações vegetais, que nesta pesquisa chegaram a um acerto médio de até 78,9%.

- A Função Discriminante de Fisher, com a implementação computacional mais simples e rápida dos três classificadores testados, mostrou-se eficiente conseguindo separar adequadamente as formações vegetais em seus diferentes estágios de regeneração, obtendo um acerto médio de 76,4% para as classes naturais. O índice de acerto médio de 65,8% obtido por este classificador para as classes antrópicas também é promissor.

- A Função Discriminante Quadrática obteve o pior desempenho entre os três classificadores estatísticos, o que pode ser atestado pelo índice de acerto médio de 68,8% obtido no sexto experimento. A análise dos resultados do sexto experimento mostrou a capacidade na separação das classes temáticas agropecuária, solo exposto e edificações (antrópicas) verificando-se um índice de acerto médio de 63,0%. Este fato é relacionado ao modelo de separação não linear adotado pela Função Discriminante Quadrática.

- A Regressão Logística com Camada Oculta obteve os melhores acertos na classificação de quase todas as classes temáticas. O índice de acerto médio de 75,9% comprova este fato. As exceções foram verificadas nas classes corpos d'água

e várzea, que têm características muito próximas sendo constatadas confusões de até 41% entre essas duas classes. O método de Newton Raphson empregado na determinação dos estimadores das funções de verossimilhança apresentou convergência para todas as classes, revelando a sua aplicabilidade para o sétimo experimento. Destaca-se ainda a facilidade de implementação do método iterativo de Newton Raphson, com exemplos de seu desenvolvimento na literatura consultada.

- Ao considerar-se os tempos de processamento necessário a obtenção da classificação da cobertura do solo é evidente que a Função Discriminante Linear de Fisher apresenta grande vantagem sendo sua execução computacional 2 vezes mais rápida que a Regressão Logística com Camada Oculta.

- Cita-se como desvantagem do procedimento de classificação da cobertura do solo apresentado a necessidade da obtenção de amostras a campo, o que pode ser minimizado através da cuidadosa fotointerpretação do local. Contudo, para obter uma separação adequada das classes temáticas faz-se necessário o uso de imagens com resolução geométrica compatível aos trabalhos almejados, o que pode constituir custos adicionais ao processo de classificação dos dados LASER.

- Finalmente, conclui-se que os objetivos deste trabalho foram todos alcançados, destacando-se o sucesso na separação das classes temáticas relativas à cobertura do solo, especialmente, aquelas relacionadas aos estágios de sucessão vegetal. Como classificador estatístico mais adequado a separação das classes temáticas aconselha-se o uso da Regressão Logística com Camada Oculta por independe da configuração dos dados, especialmente, a condição de matrizes de co-variâncias iguais para as classes temáticas. O tempo de processamento necessário a classificação quando emprega-se a Regressão Logística com Camada Oculta poderá ser diminuído através da otimização do algoritmo implementado nesta pesquisa.

Visando a continuidade de pesquisas relacionadas a este trabalho recomenda-se:

- O aprimoramento do filtro morfológico através da adoção do ajuste de planos baseados em células de grade regular com tamanho variável e mesmo a combinação com outros métodos de filtragem matemáticos já implementados;



- Realizar a comparação dos resultados obtidos através do filtro morfológico adaptado com levantamentos de campo e com precisões compatíveis com as precisões do *laser scanner*;
- Combinar o método de segmentação de imagens aos classificadores apresentados nesta pesquisa.
- Empregar as ferramentas da classificação orientada a objetos, citando como exemplos a textura e compacidade, como cofatores para os classificadores testados neste trabalho;
- Implementar um classificador de dados LASER baseado em Regressão Logística por Componentes Principais.

## REFERÊNCIAS

ALBERT, A., ANDERSON, J.A. . On The Existence Of Maximum Likelihood Estimates In Logistic Regression Models. **Biometrika**, vol. 71, 1, pp. 1-10, 1984.

ALBERTI, G. A.; da SILVA, E. A. Geração de um programa para processamento morfológico de imagens visando auxiliar o processo de extração de feições cartográficas de imagens orbitais. In: XIV Simpósio Brasileiro de Sensoriamento Remoto, Natal, 2009. **Anais**.

ANDRADE, J. B. **Fotogrametria**. Curitiba: SBEE, 2003, 2<sup>a</sup> ed.

ANDREWS, D.S.. **Simplifying Terrain Models an Measuring Terrain Model Accuracy**. Dissertação de Mestrado, Departamento de Ciências da Computação, Universidade de *British Columbia*, Canadá, 1996.

ARAKI, H.. **Fusão de Informações Espectrais, Altimétricas e de Dados Auxiliares na Classificação de Imagens de Alta Resolução Espacial**. Tese de Doutorado, Curso de Pós-Graduação em Ciências Geodésicas, Universidade Federal do Paraná, Curitiba, 2005.

AXELSSON, P.. Processing of LASER scanner data – algorithms and applications. **Journal of Photogrammetry and Remote Sensing**. Vol. 54, p. 138-147, 1999.

BAATZ, M.; BENZ, U.; DEGHANI, S.; HEYNEN, M.; HÖLTJE, A.; HOFMANN, P.; LINGENFELDER, I.; MIMLER, M.; SOHLBACH, M.; WEBER, M.; WILLHAUCK, G. **e-Cognition Professional User Guide 4**. München, Germany: Definiens Imaging, p. 486, 2004.

BALTSAVIAS, E. P. (a). A comparison between photogrammetry and LASER scanning. **Journal of Photogrammetry and Remote Sensing**. Vol. 54, p. 83-94, 1999.

BALTSAVIAS, E. P. (b). Airborne LASER scanning: basic relations and formulas. **Journal of Photogrammetry and Remote Sensing**. Vol. 54, p. 199-214, 1999.

BERNARDES, T.; FONSECA, L. M. G; MOREIRA, M. A.; CAMARGO, F. F. Estruturação de redes semânticas na classificação orientada a objeto de imagens

orbitais para mapeamento de áreas cafeeiras. In: XIV Simpósio Brasileiro de Sensoriamento Remoto. Natal, 2009. **Anais**.

BLAIR, B. J.; RABINE, D. L.; HOFTON, M. A. The LASER Vegetation Imaging Sensor: a médium-altitude, digitisation-only, airborne LASER altimeter for mapping vegetation and topography. **Journal of Photogrammetry and Remote Sensing**. Vol. 54, p. 115-122, 1999.

BRASIL, Conselho Nacional do Meio Ambiente (CONAMA). **Resolução nº 004, de 04 de maio de 1994**. Disponível em <<http://www.mma.gov.br/prot/conama/res/res94/res0494.html>>. Acesso em 02 de fevereiro de 2007.

BORTOLOT, Z.; WYNNE, R.H. Estimating forest biomass using small footprint LIDAR data: An individual tree-bases approach that incorporates training data. **Journal of Photogrammetry and Remote Sensing**. Vol. 59, p. 342-360, 2005.

BOTELHO, M. F.; SILVA, C. R. da; SHOENINGER, E. R.; CENTENO, J. A. S. Comparação dos resultados de interpoladores “Vizinho mais próximo” e “Inverso de uma distância” no cálculo de volume a partir de dados do laser scanner. In: XIV Simpósio Brasileiro de Sensoriamento Remoto. Natal, 2005. **Anais**.

BOTELHO, M. F. **Modelagem Tridimensional de Edificações Usando Dados do Sistema Laser Scanner e Imagem Orbital de Alta Resolução Espacial**. Tese (Doutorado em Ciências Geodésicas). Curso de Pós-Graduação em Ciências Geodésicas, Universidade Federal do Paraná, Curitiba, 2007.

CAMPOS, I. O. **Análise de Técnicas para o Estabelecimento de Modelos Digitais de Terrenos – DTM**. 1994. Dissertação (Mestrado em Engenharia) - Escola de Engenharia, Universidade Estadual Paulista, São Carlos, 1994.

CENTENO, J. A. S. **Sensoriamento Remoto e Processamento de Imagens Digitais**. Curitiba, Ed. C.P.G.C.G. – UFPR, 2004.

CRHISTENSEN, R. **Log-linear models and logistic regression**. Springer, New York, 1997.

CULLEN JUNIOR, L., PÁDUA, C. V. & RUDRAN, R. **Métodos de estudos em biologia da conservação e manejo da vida silvestre**. Curitiba: Ed. UFPR/ Fundação O Boticário de Proteção à Natureza, p. 539-553. 667p, 2003.

DALMOLIN, Q.; SANTOS, D. R. DOS. **Sistema LASER scanning**: conceitos e princípios de funcionamento. Curitiba: Editora da UFPR, 2003.

DISPERATI, A. A.; AMARAL, R. F. DO; SCHULER, A. B. **Fotografias aéreas de pequeno formato: aplicações ambientais**. Guarapuava: Ed. Unicentro, 2007;

DLUGOSZ, F. L. ; ROSOT, N. C.; IURK, M. C. Avaliação quantitativa da segmentação por crescimento de regiões para a discriminação de tipologias da Floresta Ombrófila Mista em imagens Ikonos. In: XIII Simpósio Brasileiro de Sensoriamento Remoto, Florianópolis, Brasil, INPE, p. 1639-1646, 2007. **Anais**.

ESRI. **ArcInfo Workstation Manual**. New York, 2005.

FELGUEIRAS, C. A.. **Modelagem Numérica de Terreno**. In: CÂMARA, Gilberto; DAVIS, Clodoveu; MONTEIRO, Antônio Miguel Vieira (Ed.). **Introdução à ciência da geoinformação**. São José dos Campos: INPE, 2001. p. 38. Publicado como: INPE-8081-PRE/3896. Disponível em: <<http://mtc-m12.sid.inpe.br/rep-sid.inpe.br/sergio/2004/04.19.14.52>>. Acesso em: 14 abril 2009.

GEMAEL, C., **Introdução ao ajustamento de observações: Aplicações geodésicas**. Curitiba, Editora da UFPR, 1994.

GUIMARÃES, I. A. **Modelos de Regressão Logística Oculto e de Componentes Principais para Reconhecimento e Classificação de Padrões com Variável Resposta Politômica**. Tese (Doutorado em Métodos Numéricos) – Setor de Tecnologia, Universidade Federal do Paraná. Curitiba, 2006.

HABIB, A. F.; GHANMA, M.; MORGAN, M.; MITISHITA, E. **Integration of LASER and Photogrammetric Data for Calibration Purposes**. In: XX ISPRS Congress. Istanbul, 2004. **Anais**.

HOSMER, D. W.; LEMESHOW, S. **Applied logistic regression**. New York: Wiley Interscience, 1989;

HYPPA, J.; PYYSALO, U.; HYPPA, H.; SAMBERG, A. Elevation Accuracy of LASER Scanning-Derived Digital Terrain and Target Models in Forest Environment. In: EARSeL-SIG-Workshop LIDAR. Dresden, 2000. **Proceedings**.

JOHNSON, D. E. **Applied multivariate methods for data analysts**. U.S.A.: Duxbury Press, 1998.

KOBLER, A.; PFEIFER, N.; OGRINC, P.; TODOROVSKI, L.; OSTIR, K.; DZEROSKI, S. Repetitive interpolation: A robust algorithm for DTM generation from Aerial LASER Scanner Data in Forest Terrain. **Remote Sensing of Environment**, Vol. 108, pp 9-23, 2007.

KIM, S.; MCGAUGHEY, R. J.; ANDERSEN, H. E.; SCHREUDER, G. Tree species differentiation using intensity data derived from leaf-on and leaf-off airborne laser scanner data. **Remote Sensing of Environment**, V. 113, pp 1575-1586, 2009.

LEE, A.; LUCAS, R.; BRACK, C. Quantifying vertical forest stand structure using small footprint lidar to assess potential stand dynamics. In: ISPRS working group VIII/2 - LASER-Scanners for Forest and Landscape Assessment. Freiburg, 2004. **Proceedings**.

LOCH, C.. **A Interpretação de Imagens Aéreas**. Florianópolis: Editora da UFSC, 1993.

MACHADO, A. M. L. **Extração Automática de Contornos de Edificações Utilizando Imagem Gerada por Câmara Digital de Pequeno Formato e Dados LIDAR**. Tese (Doutorado em Ciências Geodésicas). Curso de Pós-Graduação em Ciências Geodésicas, Universidade Federal do Paraná, Curitiba, 2006.

MINGOTI, S. A. **Análise de Dados Através de Estatística Multivariada**. Belo Horizonte: Editora da UFMG, 2005.

MITISHITA, E. A. **Monorestituição Digital de Aerofotos, Associada com Sistema de Computação Gráfica C.A.D., para fins de Mapeamento na Área Florestal**. Tese (Doutorado em Engenharia Florestal) – Setor de Ciências Agrárias, Universidade Federal do Paraná. Curitiba, 1997.

MORSDORF, F.; KOETZ, B.; MEIER, E.; ITTEN, K. J.; ALLGOWER, B. The Potencial of Discrete Return, Small Footprint Airborne Laser Scanning Data For Vegetation Density Estimation. ISPRS WG III/3 – Workshop LASER scanning . Enschede-Holanda, 2005. **Proceedings**.

MOFFIET, T.; MERNGERSE, C.; WITTE, C.; KING, R.; DENHAM, R.. Airborne LASER scanning: Exploratory data analysis indicates potential variables for

classification of individual trees or forest stands according to species. **Journal of Photogrammetry and Remote Sensing**. Vol. 59, p. 115-122, 2005.

NAMIKAWA, L. M.; FELGUEIRAS, C. A.; MURA, J. C.; ROSIM, S.; LOPES, E. S. S.. **Modelagem Numérica de Terreno e Aplicações**. São José dos Campos, 2003. Disponível em: <<http://mtc-m12.sid.inpe.br/rep-/sid.inpe.br/marciana/2003/03.10.11.36>>. Acesso em 12 de abril de 2009.

NARDINOCCHI, C.; FORLANI, G.; ZINGARETTI, P. Classification and filtering of LASER data. **International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences**, Vol. XXXIV, 3/W13, 2003.

ORKA, H. O.; NAESSET, E.; BOLLANDSÅS, O. M. Utilizing airborne LASER intensity for tree species classification. In: Workshop 'LASER Scanning 2007 and SilviLASER 2007'. Espoo, Finland. Vol. XXXVI, pp 303-304, 2007. **Proceedings**.

ORKA, H. O.; NAESSET, E.; BOLLANDSÅS, O. M. Classifying species of individual trees by intensity and structure features derived from airborne LASER scanner data. **Remote Sensing of Environment** Vol. 113, pp 1163-1174, 2009.

PAIVA, Y. G.; SILVA, K. R.; PEZZOPANEL, J. E. M.; QUINTÃO, A. A.; XAVIER, A. C.; CECÍLIO, R. A. Estudo de sítios florestais e fragmentos de Mata Atlântica na Bacia do Rio Itapemirim usando imagens CBERS e dados SRTM. In: XIII Simpósio Brasileiro de Sensoriamento Remoto, Florianópolis, Brasil, INPE, p. 1793-1800, 2007. **Anais**.

PAIXÃO, L. A. **Avaliação da Qualidade do Óleo Isolante em Transformadores com o Emprego da Função Discriminante Quadrática**. Dissertação. Curso de Pós-Graduação em Métodos Numéricos em Engenharia, Universidade Federal do Paraná, Curitiba, 2006.

PFEIFER, N.; GORTE B.; ELBERINK, S. O. Influences of Vegetation on LASER Altimetry - Analysis and Correction Approache. ISPRS WG VIII/2 - Workshop LASER-Scanners for Forestand Landscape Assessment. Freiburg, Germany, 2004. **Proceedings**.

RODERJAN, C. V.; Y. S. KUNIYOSHI & F. GALVÃO. As regiões Fito-geográficas do Estado do Paraná. **Acta Florestalia Brasiliensis**, Curitiba, 1993.

ROUSSEEUW, P. J.; CHRISTMANN, A. Robustness against separation and outliers in logistic regression. **Computational Statistic and Data Analysis**. Vol. 43, p 315-332, 2003.

SITHOLE, G.; VOSSELMAN G.. Experimental Comparison of Filter Algorithms for Bare-Earth Extraction from Airborne LASER Scanning Point Clouds. **Journal of Photogrammetry&Remote Sensing**, Vol. 59, pp 85-101,2004.

SITHOLE, G. Segmentation and Classification of Airborne LASER Scanner Data. TU Delft, **Publications on Geodesy of the Netherlands Commission of Geodesy**, Vol. 59, 2005.

TENG, W. L. **Fundamentals of Photographic Interpretations**. In: Manual of Photographic Interpretation. 2<sup>a</sup> ed., Maryland: ASPRS, 1997.

VELOSO, H. P.; FILHO, A. L. R. R.;LIMA, J. C. A.. **Classificação da vegetação brasileira, adaptada a um sistema universal**. Rio de Janeiro: IBGE, 1991.

WEHR, A.; LOHR, U. Airborne LASER scanning – an introduction and overview. **Journal of Photogrammetry and Remote Sensing**. Vol. 54 (1), p. 68-82, 1998.

WOLF, P. R.; DEWITT, B. A. **Elements of Photogrammetry: with applications in GIS**. 3<sup>a</sup> ed. U.S.A: MacGraw - Hill Book Company, 2000.

ZAK, M. R.; CABIDO, M. Spatial patterns of the Chaco vegetation of central Argentina: integration of Remote Sensing and Phytosociology. **Applied Vegetation Science**, v. 5, p. 213-226, 2002.

ZAKSEK, K.; PFEIFER, N.. An Improved Morphological Filter for Selecting Relief Points from a LIDAR Point Cloud in Steep Areas with Dense Vegetation. **Technical Report** performed in 2006, Delft Institute of Earth Observations.

## APÊNDICE

### COEFICIENTES DAS FUNÇÕES DISCRIMINANTES QUADRÁTICAS

A metodologia para classificação das células da grade LASER empregando a Função Discriminante Quadrática foi proposta por PAIXÃO(2006) em sua dissertação de mestrado. O cálculo da probabilidade da classe temática é implementado de forma matricial:

$$Pr_i = \frac{1}{2} Q_i L_i C_i$$

Sendo:

$Pr_i$  - Probabilidade de a célula selecionada pertencer a classe  $i$ ;

$Q_i$  - Coeficientes de segundo grau para a classe  $i$ ;

$L_i$  - Coeficientes de primeiro para a classe  $i$ ;

$C_i$  - Constante relativa a classe  $i$ ;

$i = 1, 2, \dots, 12$ .

A partir do ajuste das 12 funções de discriminação foram obtidas as seguintes matrizes:

#### 1ª Função de discriminação:

Coeficientes de Segundo Grau:

-0,0555	0,06141	0,10971	0,06279	0,00936	0,51264
0	-0,0367	-0,1137	-0,0622	-0,015	0,13063
0	0	-0,6113	0,42807	-0,0063	1,53563
0	0	0	-1,5934	0,25858	-1,3456
0	0	0	0	-0,0219	0,0431
0	0	0	0	0	-5,9163



Coeficientes de Primeiro Grau:

79,127	-39,385	-76,124	-40,25	-5,1563	-401,54
--------	---------	---------	--------	---------	---------

Constante: -28494,245559.

## 2ª Função de discriminação:

Coeficientes de Segundo Grau:

-0,1693978	-0,0422502	-0,2948504	-1,9929129	0,0550966	4,0960308
0	-0,0506869	0,0213916	-0,8174761	-0,0773152	1,3368280
0	0	-0,4690870	-0,6844426	0,0566734	6,1304352
0	0	0	-18,5204926	1,1463883	23,1485449
0	0	0	0	-0,1690223	0,5546965
0	0	0	0	0,0000000	-194,4483851

Coeficientes de Primeiro Grau:

319,76489	50,21592	273,94785	1938,76905	-38,84073	-3468,57667
-----------	----------	-----------	------------	-----------	-------------

Constante: -151896,6967.

## 3ª Função de discriminação:

Coeficientes de Segundo Grau:

-0,018560	-0,025487	-0,029099	-0,266557	0,041024	1,497122
0	-0,024131	-0,197877	-0,122846	0,042674	4,027109
0	0	-0,607625	0,582999	0,092175	19,272603
0	0	0	-3,295215	0,605734	-1,391220
0	0	0	0	-0,045962	-2,608866
0	0	0	0	0	-270,162688

Coeficientes de Primeiro Grau:

32,672231	21,651496	19,943847	244,537353	-36,041332	-991,347126
-----------	-----------	-----------	------------	------------	-------------

Constante: -14564,25.

**4ª Função de discriminação:**

Coeficientes de Segundo Grau:

-2,67938	-0,15355	7,11845	-2,69941	1,84041	-7,54225
0	-0,01813	0,13656	-0,16466	0,04809	-0,34594
0	0	-5,05323	3,61636	-2,36000	9,51477
0	0	0	-1,11585	0,99702	-3,95669
0	0	0	0	-0,41874	2,69685
0	0	0	0	0	-6,89941

Coeficientes de Primeiro Grau:

4024,13601	118,36363	-5340,48760	2034,34660	-1377,42848	5677,43853
------------	-----------	-------------	------------	-------------	------------

Constante: -1511151,953.

**5ª Função de discriminação:**

Coeficientes de Segundo Grau:

-7,96189E+15	-1,23625E+15	8,25395E+15	-1,53675E+16	1,61298E+15	-7,96189E+15
0	-3,37343E+14	-3,23501E+15	-7,08360E+15	4,63427E+14	0
0	0	-1,01722E+16	-4,77529E+16	1,26993E+15	0
0	0	0	-2,40177E+16	5,26069E+15	0
0	0	0	0	-1,79238E+14	0
0	0	0	0	0	0

Coeficientes de Primeiro Grau:

1,19771E+19	9,30959E+17	-6,25277E+18	1,16555E+19	-1,21302E+18	-4,76231E+19
-------------	-------------	--------------	-------------	--------------	--------------

Constante: -4,50E+21.

**6ª Função de discriminação:**

Coeficientes de Segundo Grau:

-2,07507	-0,25411	5,74965	-12,58006	0,11788	-2,46687
0	-0,01108	0,24500	-0,74651	0,01078	-0,45982
0	0	-6,69572	25,63167	0,03673	13,62577
0	0	0	-54,83716	-0,34716	-55,99641
0	0	0	0	-0,07323	-0,71698
0	0	0	0	0	-46,36441

Coeficientes de Primeiro Grau:

3121,52350	191,44356	-4361,16264	9589,27249	-84,99766	2070,20866
------------	-----------	-------------	------------	-----------	------------

Constante: -1174209,587.

### 7ª Função de discriminação:

Coeficientes de Segundo Grau:

-62,13549	4,28210	-66,50638	91,87941	-8,33190	-511,95151
0	-0,08353	3,58291	-3,89278	0,35347	19,03153
0	0	-71,40170	135,07066	-9,81647	-377,28177
0	0	0	-89,18067	10,59822	430,89173
0	0	0	0	-0,42884	-39,62551
0	0	0	0	0	-1169,00210

Coeficientes de Primeiro Grau:

93060,80894	-3200,13126	49256,68005	-68338,45209	6212,64192	383010,06182
-------------	-------------	-------------	--------------	------------	--------------

Constante: -34845973,04.

### 8ª Função de discriminação:

Coeficientes de Segundo Grau:

-0,01071	0,01554	-0,02141	-0,02335	0,02071	-0,30485
0	-0,01000	0,00979	0,00900	-0,01821	0,07846
0	0	-0,04612	-0,06658	0,00781	-0,13988
0	0	0	-0,96394	0,09167	-2,53707
0	0	0	0	-0,05612	-0,11265
0	0	0	0	0	-8,22365

Coeficientes de Primeiro Grau:

17,87964	-11,58340	19,36024	28,45112	-12,84847	314,27909
----------	-----------	----------	----------	-----------	-----------

Constante: -7705,658098.

**9ª Função de discriminação:**

Coeficientes de Segundo Grau:

-0,02056	-0,00022	0,08623	-0,04465	-0,03517	0,19928
0	-0,01808	0,04258	0,03366	-0,00134	1,94052
0	0	-0,12574	0,07176	0,09014	-1,98839
0	0	0	-0,59635	-0,00452	-2,71399
0	0	0	0	-0,07117	0,48064
0	0	0	0	0	-106,55004

Coeficientes de Primeiro Grau:

35,03096	0,33786	-74,63155	40,94733	33,26640	-60,33238
----------	---------	-----------	----------	----------	-----------

Constante: -15058,07559.

**10ª Função de discriminação:**

Coeficientes de Segundo Grau:

-0,59694	-0,10132	-1,60880	2,47754	0,40274	37,27340
0	-0,01693	-0,06826	-0,27167	0,11036	6,62392
0	0	-1,90790	7,35734	0,16680	2,55105
0	0	0	-12,16635	1,38282	63,95767
0	0	0	0	-0,24578	-28,71267
0	0	0	0	0	-1399,92939

Coeficientes de Primeiro Grau:

967,60822	73,21224	1437,07194	-2400,85581	-280,29550	-25668,03196
-----------	----------	------------	-------------	------------	--------------

Constante: -398477,3439.

**11ª Função de discriminação:**

Coeficientes de Segundo Grau:

-24,56623	-6,38498	46,49068	-88,22180	10,27025	47,86339
0	-0,56322	4,42842	-13,27562	1,56449	6,61303
0	0	-28,83583	71,82179	-8,69173	-48,38250
0	0	0	-88,94420	20,43549	87,88520
0	0	0	0	-1,22098	-10,65364
0	0	0	0	0	-28,37821

Coeficientes de Primeiro Grau:

45634,82297	5957,25943	-42994,55126	82107,74773	-9555,58772	-44437,25637
-------------	------------	--------------	-------------	-------------	--------------

Constante: -21194513,71.

## 12ª Função de discriminação:

Coeficientes de Segundo Grau:

-1,75481	0,01367	0,22098	1,41141	0,09963	-3,08109
0	-0,00421	0,00362	-0,04689	-0,00590	-0,10719
0	0	-0,09509	-0,08040	0,01474	0,36881
0	0	0	-0,94087	-0,01008	-6,61078
0	0	0	0	-0,01927	0,28238
0	0	0	0	0	-106,32950

Coeficientes de Primeiro Grau:

2636,24991	-9,17140	-166,05503	-1045,29139	-73,91958	2591,46024
------------	----------	------------	-------------	-----------	------------

Constante: -990356,3385.

## **ÓCULOS DE ANAGLIFO**

Detalhes sobre o funcionamento do método do anaglifo poderão ser encontrados em ANDRADE (2003).